Identification, characterization and engineering of glutathione-dependent, lignin-degrading enzymes

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Hauke Voß
aus Neumünster

1. Referentin: Professorin Dr. Anett Schallmey

2. Referent: Professor Dr. Dieter Jahn

eingereicht am: 21.10.2019

mündliche Prüfung (Disputation) am: 18.12.2019

Druckjahr 2020

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch die Mentorin der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

J. Husarcíková, H. Voß, P. Domínguez de María, A. Schallmey, "*Microbial* β-*etherases and glutathione lyases for lignin valorisationin biorefineries: current state and future perspectives*", *Applied Microbiology and Biotechnology*, Vol 102, 5391–5401, (**2018**).

H. Voß, C. A. Heck, M. Schallmey, A. Schallmey, "*Database mining for novel bacterial* β-*etherases, glutathione-dependent lignin-degrading enzymes*", Applied and Environmental Microbiology, Vol 86, No 2, (**2020**).

Posterbeiträge

H. Voß, P. Picart, C. Diederich, A. Schallmey, "*Application of* β-*etherases and glutathione lyases in lignin valorization*", **Biocat 2016**, 8th International Congress of Biocatalysis, Hamburg, Germany, (**2016**).

H. Voß, C.Diederich, A. Schallmey, "*PROTEIN ENGINEERING OF LIG-TD, A GLUTATHTION LYASE*"**, BioTrans 2017**, 13th International Symposium on Biocatalysis and Biotransformation, Budapest, Hungary, (**2017**).

H. Voß, C. A. Heck, A. Schallmey, "*IDENTIFICATION OF NEW* β-*ETHERASES BY PEPTIDE PATTERN RECOGNITION*", **ICRC 2018**, International CeBiTec Research Conference 2018 Reaction Concepts for Industrial Biocatalysis, Bielefeld, Germany, (**2018**).

H. Voß, C. A. Heck, A. Schallmey, "*IDENTIFICATION OF NEW* β-*ETHERASES BY PEPTIDE PATTERN RECOGNITION*", **Biocat 2018**, 9th International Congress of Biocatalysis, Hamburg, Germany, (**2018**).

H. Voß, C. A. Heck, A. Schallmey, "*IDENTIFICATION OF NEW* β-*ETHERASES BY PEPTIDE PATTERN RECOGNITION*", **Procompas PhD Symposium 2018**, Braunschweig, Germany, (**2018**).

Table of contents

Abstract

Lignin is one of the most abundant polymers in nature. Due to this it is a very interesting source for the provision of aromatic platform chemicals and biofuels. Until now, however, lignin is not industrially used since the required lignin depolymerization is very challenging. β-Etherases and glutathione lyases are enzymes with proven potential in depolymerizing lignin and one possible solution for this problem.

However, the number of these enzymes is still rather limited. Using classical public database mining in combination with the peptide pattern recognition algorithm, 96 putatively novel β-etherases have been identified in this thesis, some even originating from bacteria outside the order *Sphingomonadales*. Of these, a subset of 13 enzymes was chosen for further biochemical characterization. All tested enzymes are active β-etherases and some enzymes displayed up to three-fold higher activity compared to the previously known β-etherases. Furthermore, the sequence diversity within β-etherase family was increased and detailed information about conserved sequence areas of β-etherases was generated. In combination with performed structural analyses, this is a great step in understanding the β-etherase mechanism and to identify amino acids essential for their activity. Moreover, through phylogenetic analyses an interesting clustering of lignin-degrading enzymes, LigF-, NaLigF-2-type, and hetero-dimeric β-etherases, was revealed. The gained phylogenetic information will likely also lead to the identification of new types of lignin-degrading β-etherases in the future.

The second part of this thesis dealt with the optimization of glutathione lyase LigG-TD. First, a high-throughput assay for activity screening of glutathione lyases was developed. This assay is based on the NADPH-dependent, glutathione reductase-catalyzed reduction of oxidized glutathione GSSG, one product of the glutathione lyase reaction. Hence, based on the consumption of NADPH the glutathione lyase activity is indirectly measured. This assay was subsequently applied in a protein engineering campaign to increase the activity of the glutathione lyase LigG-TD. Upon screening six site-saturation mutagenesis libraries, several LigG-TD mutants with improved activity could be identified. Overall, with only one round of protein engineering the activity of LigG-TD was increased by 20%. The developed screening assay can be also applied for the protein engineering of other glutathione lyases in order to increase their activity or stability.

Zusammenfassung

Lignin ist eines der am häufigsten vorkommenden, natürlichen Polymere und eine interessante Quelle für die zukünftige Gewinnung von aromatischen Chemikalien und Biokraftstoffen. Bisher wird Lignin aufgrund seiner schwierigen Depolymerisierung noch nicht als Ausgangsstoff für neue Produkte im industriellen Maßstab genutzt. Eine mögliche Lösung für dieses Problem stellen β-Etherasen und Glutathionlyasen dar, Glutathion-abhängige Enzyme, welche die β-*O*-4 Aryletherbindungen im Lignin spezifisch spalten können.

Allerdings waren bisher nur wenige Vertreter dieser Enzymgruppen bekannt. Daher wurden im Rahmen dieser Arbeit durch klassische Datenbankensuche in Kombination mit dem *peptide pattern recognition* Algorithmus 96 neue putative β-Etherasen identifiziert, wovon 13 Enzyme biochemisch charakterisiert wurden. Dabei konnte für alle getesteten Enzyme die erwartete Aktivität nachgewiesen werden. Einige Vertreter wiesen hierbei sogar eine bis zu dreifach höhere Aktivität als die bisher bekannten β-Etherasen auf. Durch die Identifizierung der neuen β-Etherasen wurde die Sequenzdiversität der Enzymfamilie erhöht und in Verbindung mit durchgeführten strukturellen Analysen Informationen über die an der Funktion beteiligten Aminosäuren generiert. So wurde eine Vielzahl von konservierten Aminosäuren gefunden, denen eine katalytische oder strukturelle Funktion zugewiesen werden konnte. Zusätzlich wurden durch phylogenetische Analysen interessante Verwandtschaftsbeziehungen innerhalb der Enzymfamilie offengelegt, was zu der Entdeckung neuer Arten von Lignin-abbauenden β-Etherasen führen kann.

Der zweite Teil dieser Arbeit beschäftigte sich mit Optimierung der Glutathionlyase LigG-TD durch *protein engineering*. Hierfür wurde zunächst ein *high-throughput* Assay für das Aktivitätsscrenning von Glutathionlyasen entwickelt. Der Assay basiert auf der NADPH-abhängigen, Glutathionreduktase-katalysierten Reduktion von oxidiertem Glutathion GSSG, einem Produkt der Glutathionlyase-Reaktion. Daher kann basierend auf dem NADPH-Verbrauch die Glutathionlyaseaktivität indirekt bestimmt werden. Anschließend wurde der Assay zum Aktivitätssreening von sechs Sättigungsmutagenesebibliotheken genutzt. Hierbei wurden mehrere LigG-TD-Varianten mit verbesserter Aktivität identifiziert. Insgesamt wurde die Aktivität von LigG-TD mit einer Runde *protein engineering* um 20% gesteigert. Zukünftig kann dieser *high-throughput* Assay auch für das *protein engineering* anderer Glutathionlyasen eingesetzt werden, um deren Aktivität oder Stabilität zu verbessern.

1 Introduction

1.1 Renewable resources

Renewable resources such as biomass, bioethanol, wind or solar energy are becoming increasingly important as energy resources. In the year 2018, 40.6% of the electricity was generated by renewable resources in Germany.[1] Also regarding car mobility, renewable energy carriers like bioethanol or biomethane are very important. Bioethanol is mixed into normal Otto-fuel and contributed to around 6% of the total fuel market in Germany in 2018.[2] Bioethanol is typically generated out of grain, grain maize, sugar beet, and sugarcane.[3] Since these starch or sugar rich plants are also used for food and feed, bioethanol competes with the nutritional utilisation of these plants. This is also true for other chemicals produced based on renewable resources. Industrial processes are running for the production of different basic chemicals such as propylene, ethylene glycol, and acetic acid using ethanol or glucose as feedstock.[4]

Lignocellulose is a possible alternative renewable resource to overcome this drawback of the starch-based feedstock for example in the production of bioethanol. Lignin, as one part of lignocellulose, was also tested as precursor for carbon fibres. But until now, these processes are not able to compete with traditional ones, economically.[5]

1.2 Lignin

Lignin forms together with cellulose and hemicellulose lignocellulose, the major component of the plant cell walls. The natural function of lignocellulose is to provide stability and structure to the cell wall. Cellulose forms long linear fibers like steel strives for tensile and flexural strength, while lignin fills the gaps between the cellulose fibres and provides rigidity (Figure 1.1).[6,7]

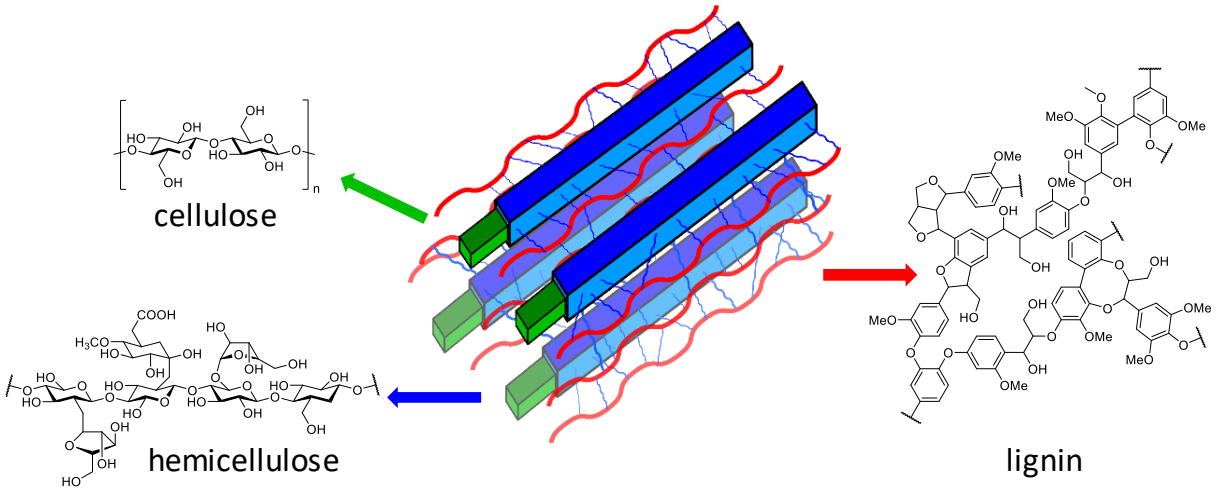


Figure 1.1 Scheme of lignocellulose structure with cellulose in green, hemicellulose in blue, and lignin in red. Adapted from Zakzeski et al.[8]

The lignin content in dry softwood is about 26-32%, whereas hardwood contains 20-25%.[9] The omnipresence of lignin in lignified plants is the big advantage of lignin, since its presence is not restricted to food plants, as in the case of starch. In addition to the generation of biofuel, lignin is also an interesting source for chemicals, in particular aromatics due to the high content of aromatic rings in lignin.[8,10,11] That also differentiates lignin from most other biomacromolecules such as proteins, polysaccharides, and lipids, which have a rather low or no aromatic content.

Lignin is easily available in big quantities, since it occurs as a waste product in different industrial processes. For example, the pulp and paper production generates up to 130 tons of lignin per year as side product.[12] Until now, most of the lignin is burned for the generation of heat and energy.[8] The target of current research is to generate new value from the vast amounts of lignin produced. For example, the production of plastic polymers using lignin as source for the monomer production is in debate.[13] Also, the EU founded two projects called “Liberate” and “B-LigZymes” about the future usage of the lignin feedstock.[14,15]

1.3 Lignin synthesis and structure

Lignin is an aromatic polymer synthesized out of the three monolignols *p*-coumaryl alcohol (H), coniferyl alcohol (G), and sinapyl alcohol (S). The lignin monomers are coupled by oxidative radical processes in the plant cell wall.[11] Involved in this process are peroxidases, which generate radicals in the presence of H_2O_2.[11] Due to this random synthesis process, lignin does not have a defined structure and consist of various C-O and C-C bonds.[11,16]

Generally, the β-*O*-4 aryl ether bond is the most abundant one with 45-60% quantity. Other common bond types in lignin are β-β, β-5, 5-5 and 5-*O*-4.[11] The different structure elements found in lignin are shown in Figure 1.2.

The composition of G, S, and H as well as the abundance of different bond types in lignin varies among different lignin sources. Softwood lignin consists mainly of G units, while hardwood lignin consists of G and S units in nearly equal amounts. H units are nearly not present in both kinds of lignin. In contrast, grass lignin contains higher quantities of H units compared soft- and hardwood linin, but absolute amounts are still low.[17,18]

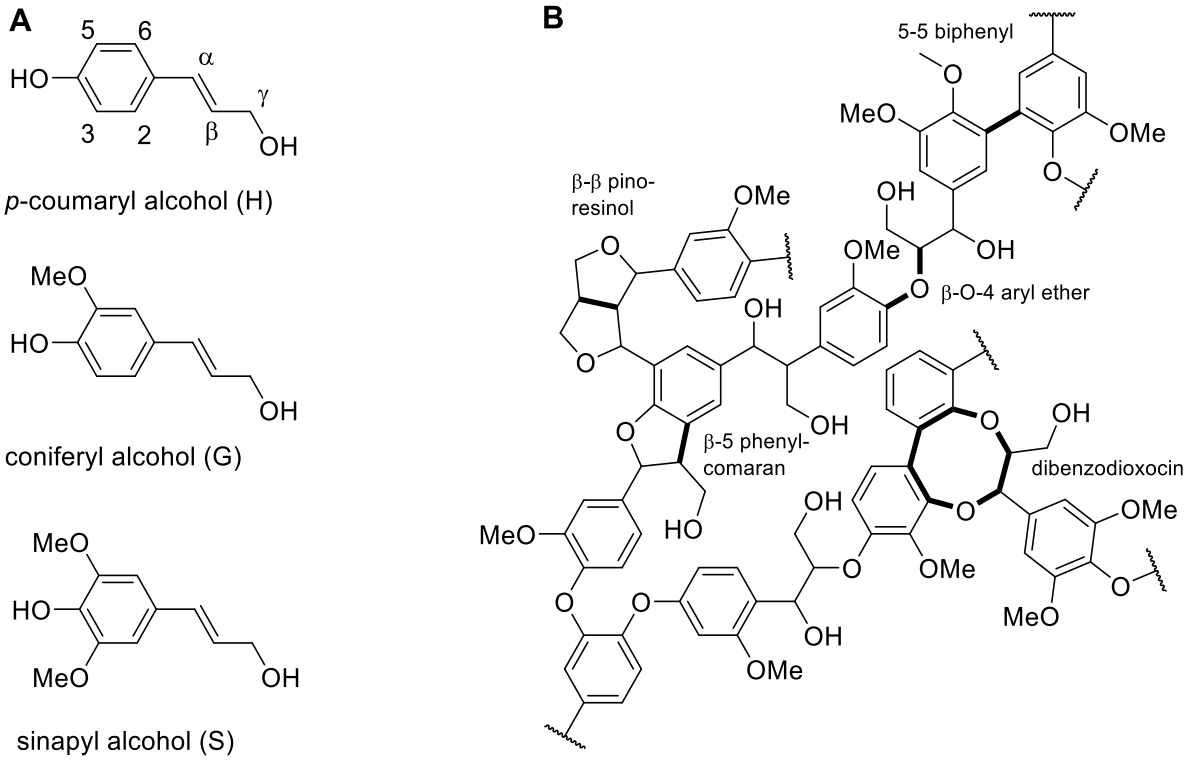


Figure 1.2 Structure of the monolignols (**A**) *p*-coumaryl alcohol (H), coniferyl alcohol (G), and sinapyl alcohol (S) as well as a representative structure of lignin (**B**). The most abundant linkages types found in lignin highlighted in bold. Adapted from Picart et al.[19]

1.4 Lignin pretreatment and isolation

Gallezot et al. indicated three general strategies for the usage of lignocellulosic biomass and lignin (Figure 1.3).[20] The first is the pyrolysis and gasification of the biomass to syngas, which can be used to synthesize biofuel and other chemicals (see also Figure 1.6). In the second and third strategy, the components of the lignocellulose are first separated by different pretreatment processes and afterwards processed individually. The aim of the second strategy is to remove the functional groups of lignin monomers to produce aromatic bulk and fine

chemicals, whereas the third strategy is to depolymerize lignin selectively by retaining the functional groups and generating directly valuable compounds such as vanillin.[8,20]

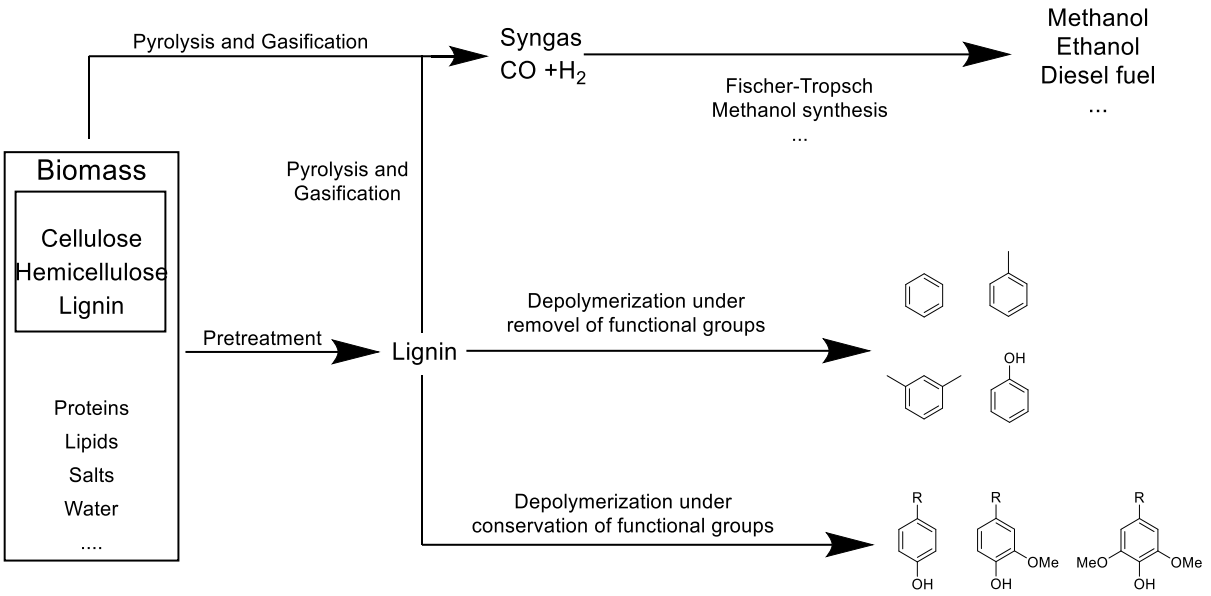


Figure 1.3 Strategies for the usage of biomass and lignin for the production of biofuel, fine, and bulk chemicals. Adapted from Zakzeski et al.[8]

Beside the lignin source, the different pretreatment processes have a big influence on the structure, composition, and the molecular weight of the isolated lignin. Common processes are the Kraft, the Lignosulfonate, and the Organosolv process.[8,21] In the Kraft and the Lignosulfonate process harsh conditions are used to solubilize the lignin for its separation from hemicellulose and cellulose.[22]

The Kraft lignin process is a very common lignin isolation process and the predominant process in the pulp and paper industry. In this process high temperatures ranging from 150-180 °C as well as high pH mediated by addition of sodium hydroxide and sodium sulfate are used. Due to these very harsh conditions, various linkages in natural lignin are broken or formed. This has an influence on the distribution of linkages found in Kraft lignin compared to untreated lignin. For example, the 5-5 linkage is very stable and, in some cases, even formed under the Kraft process conditions. Therefore, 5-5 linkages are enriched during the Kraft process. Also, stilbene linkages are formed and sulphur atoms are incorporated during the Kraft process, which are not present in untreated lignin (Figure 1.4).[23] Kraft lignin is soluble in alkaline solutions and in strong polar organic solvents. The average molecular weight is in the range between 1,000 and 3,000 Da, while the polydispersity is in the range from 2 to 4.[21]

Another process for the isolation of lignin is the Lignosulfonate process, which uses for example bisulfite and sulphuric acid at 180 °C.[24] In most cases, the Lignosulfonate process is operated at low pH conditions, but the process has a very broad pH working range between 2 and 12.[21] The high temperature and the sulphur reagents lead to the incorporation of sulfonate groups into lignin.[21] The average molecular weight of Ligninsulfonates, with a common range between 5,000 up to 20,000 Da and a polydispersity from 4 to 9, is higher compared to Kraft lignin. Lignosulfonates are soluble in water, polar organic solvents and amines.[21]

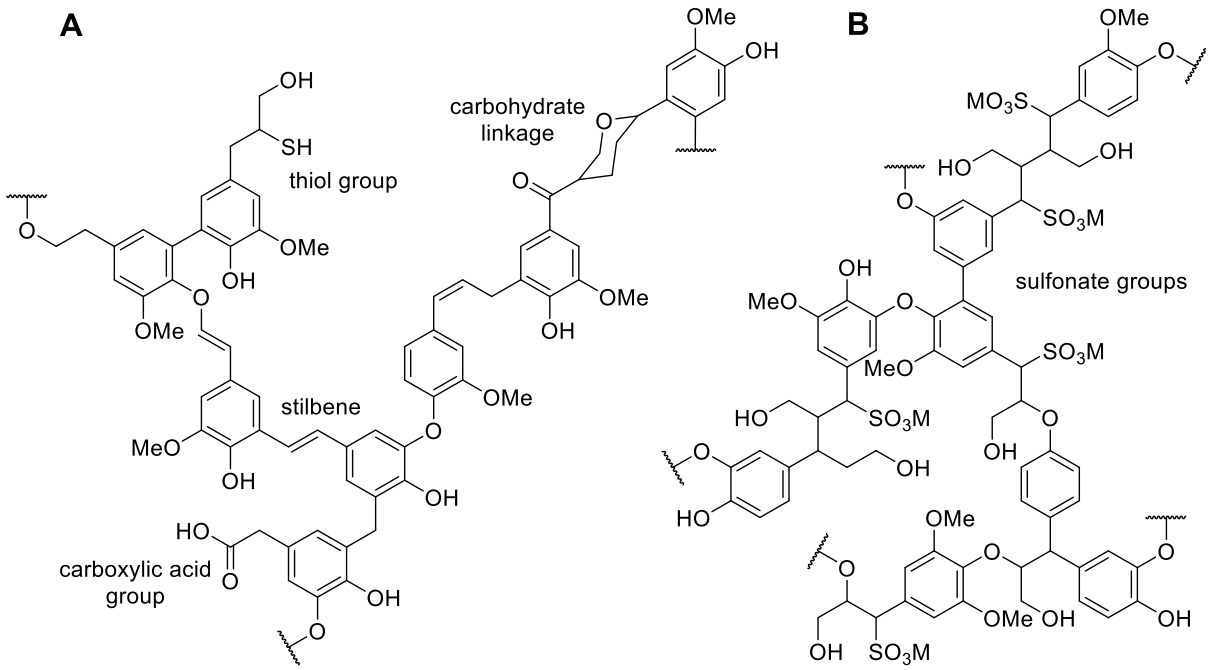


Figure 1.4 Representative structure of Kraft lignin (**A**) and Lignosulfonate lignin (**B**) with the process dependent modifications. Adapted from Zakzeski et al.[8]

Also, the solubility of the different components of lignocellulosic biomass in organic solvent/water mixtures can be utilized for pretreatment. The so called Organosolv lignin for example is generated by using a mixture of methyl isobutyl ketone (MIBK), ethanol, water, and sulphuric acid (~0.2 M), which dissolves hemicellulose and lignin at 140 °C, whereas cellulose remains in its crystalline state. After the addition of additional water, a phase separation in the liquid fraction occurs with lignin dissolved in the organic phase. The average molecular weight of Organosolv lignin is around 1,000 Da with a polydispersity from 2.4 to 6.4. Due to the rather mild extraction conditions, Organosolv lignin retains the majority of its natural structure, especially the β-*O*-4 linkage.[21] In analogy to the Organosolv process Weider et al. investigated a two-phasic system with organo-catalytic hydrolysis of the hemicellulose

fraction. A mixture of 2-methyl-tetrahydrofuran (2-MTHF) and water with 2,5-furandicarboxylic acid (FDCA) as organo-catalyst was used in this process. While cellulose remains solid, hemicellulose is hydrolysed in the water phase into sugar monomers and lignin is dissolved in 2-MTHF. Through the recycling of FDCA and 2-MTHF, this system has the potential for a very sustainable and economic process.[25]

The efficient and selective hydroxylation of glucoside bonds by enzymes is also used for the pretreatment of lignocellulose. Cellulolytic enzyme lignin (CEL) and enzymatic mild acidolysis lignin (EMAL) are produced in processes, which use cellulolytic enzymes for the hydrolysis of cellulose and hemicellulose.[26,27] The residual solid, which is mainly lignin, is dissolved in organic solvent/water mixtures, and the lignin is subsequently precipitated by an addition of additional water. Due to the presence of carbohydrate impurities, the CEL process was improved by the additional acidolysis, resulting in the EMAL process, which yields lignin of higher quality. The process is very mild and, as for the Organosolv lignin process, the number of pretreatment-induced modifications in the final lignin is very low. The average molecular weight is rather high with 17,000−30,000 Da for the CEL process and 30,000−63,000 Da for the EMAL process.[28,29]

1.5 Lignin model compounds

Lignin is a complex and heterogeneous polymer without defined structure and therefore differentiated in the different types of lignin's (section 1.3). The lignin source has a big influence on the ratio of monolignols used in the lignin synthesis, while the isolation process can introduce additional modifications. Therefore, the work with lignin is challenging and only comparable if the same kind of lignin is used.[8]

Various lignin model compounds, small organic molecules with a defined structure and usually representing only one linkage type found in lignin, were developed to tackle these problems.[8] For the characterization of lignin depolymerization reactions, chemically or biocatalytically, model compounds have multiple advantages. Reactions with lignin model compounds can be analyzed easier compared to reactions with lignin polymer, using standard methods such as HPLC (high performance liquid chromatography), LC-MS (liquid chromatography-mass spectrometry) and NMR (nuclear magnetic resonance).[30,31] Furthermore, the influence of depolymerization processes to one type of lignin linkage can be analyzed individually.

In contrast to that, lignin polymer features many different linkage types. Hence, it is much harder to determine which linkage type is cleaved during a certain depolymerization reaction. Also, the analysis methods for lignin polymer are more complex.[8] The change in lignin molecular weight due to a reaction can be analyzed by size-exclusion chromatography (SEC, also called gel permeation chromatography - GPC).[32] Sophisticated multi-dimensional NMR methods are used to analyse changes in functional groups or the abundance of linkage types in lignin as result of chemical or enzymatic reactions.[33] Small degradation products occurring upon lignin depolymerization can be analyzed using standard methods after removal of the remaining polymer.[32] But this is still challenging since the products will be very similar to each other and it is hard to predict which products will occur.

In addition, the lignin insolubility in water is a problem for the investigation of enzyme-dependent lignin degradation. Significant amounts of organic solvent need to be added to solubilize lignin, which can lead to inactivation of the biocatalyst. Lignin model compounds require less solvent addition and are therefore the more favorable compounds for studying enzymatic degradation processes.[30,34] An alternative to organic solvents in the solubilization of lignin are ionic liquids, but these can also have a negative effect to the enzyme activity.[35,36] Since the β-*O*-4 aryl ether linkage is the most abounded one found in lignin, many model compounds for this bond type have been described. A selection is shown in Figure 1.5.

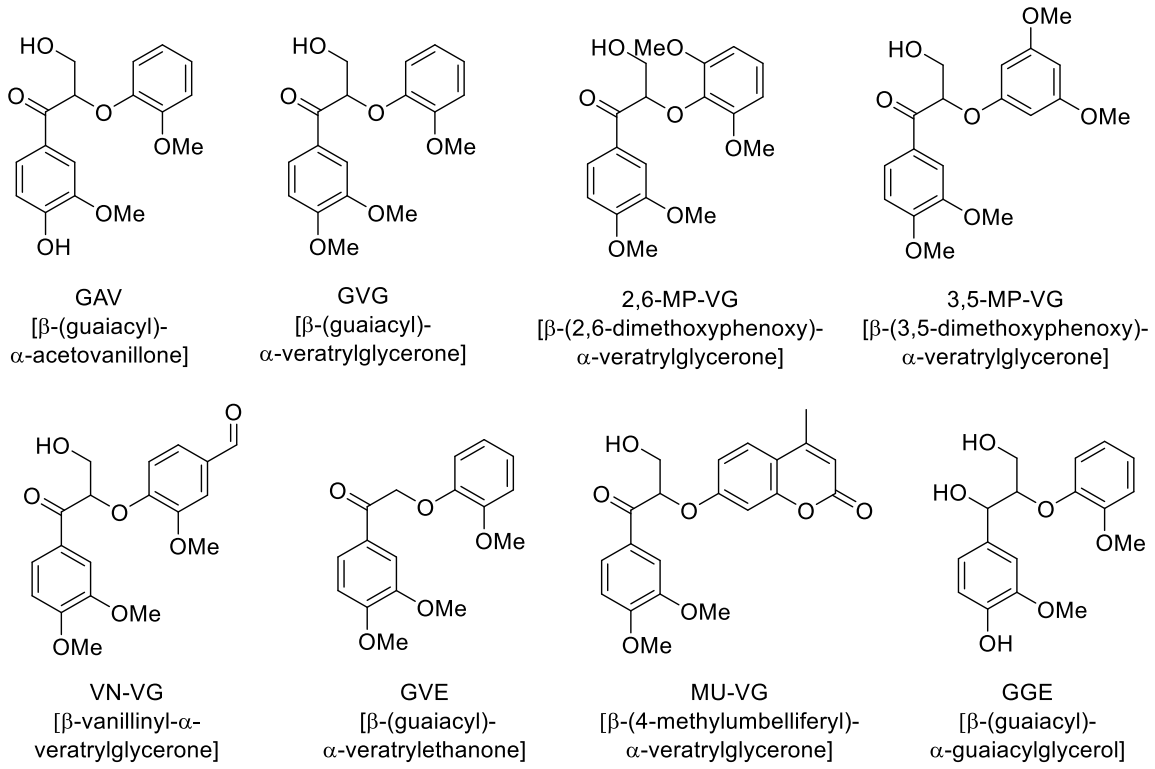


Figure 1.5 Selection of lignin model compounds, either reported in literature or used in this study, representing the β-*O*-4 aryl ether bond found in lignin.[30,31]

1.6 Lignin degradation using chemical and thermal methods

Next to the different naturally occurring enzymatic processes for lignin degradation, different chemical and thermal methods were invented by men. Thermal methods mainly refer to pyrolysis and gasification. Gasification is the total conversion of carbon-rich material, in this context biomass or lignin, to syngas under extreme heat. The process is performed at a temperature above 700 °C and a defined concentration of oxygen (mostly low concentrated) and steam.[37] The resulting syngas is a mixture of hydrogen, carbon monoxide, and carbon dioxide. The components of syngas are used to produce different chemicals such as methanol, and alkanes of different length, which can be used as diesel fuel. Furthermore, the hydrogen of syngas is used for ammonia production in the Haber Bosch process (Figure 1.6).[37]

Pyrolysis occurs under less but still extreme conditions. Depending on the heat, the frequency of linkages in lignin differs. Under controlled conditions, pyrolysis can be either used for the isolation of lignin from lignocellulose as pyrolysis lignin, or for the production of lignin monomers out of the lignin polymer. For the depolymerization, typically temperatures between 400 °C and 600 °C are used.[38] As product of lignin pyrolysis an oil containing aromatic monomers with a yield up to 54 wt.% is obtained.[39]

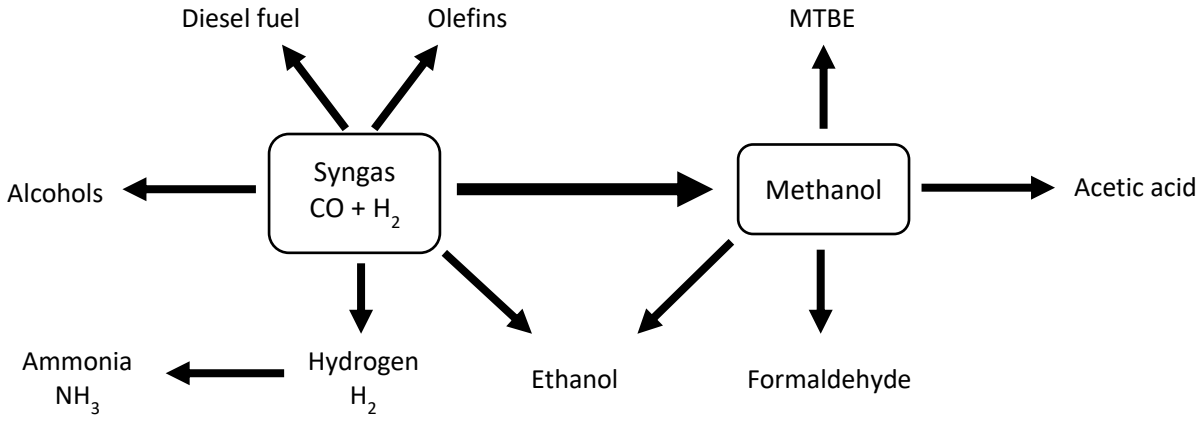


Figure 1.6 Selection of possible chemical products synthesized from syngas.

Chemical depolymerization of lignin is in general performed using rather harsh conditions, i.e. high temperatures and/or very reactive reagents. Many oxidative and reductive processes use oxygen or hydrogen, respectively, at high temperatures.[28] Anastas and co-workers developed a copper-catalyzed hydrogenation at 180 °C,[40] whereas Westwood and co-workers reported a process using lower temperatures. In this latter process, lignin is first catalytically oxidized using DDQ and afterwards cleaved by zinc ions at 80 °C.[33] The yield of lignin monomers in the process of Anastas and co-workers was excellent with around 73%. Nevertheless, it is questionable if under economic and sustainability aspects these processes are useful for the valorization of lignin, due to higher temperatures and the usage of heavy metal catalysts.[28] Further information about chemical lignin valorization is given in the review by Sun et al. and Rinaldi et al.[12,28]

1.7 Lignin-degrading enzymes

Lignin-degrading enzymes are found in many fungi and bacteria. In general, they can be grouped in mediator-dependent and mediator-independent enzymes. Mediator-dependent enzymes are not able to directly bind lignin polymer in their active sites and hence, require small molecules, so called mediators, for the degradation of the lignin polymer. Prominent examples for mediator-dependent enzymes are lignin peroxidases (LiPs), manganese peroxidases (MnPs), and laccases.[41] The glutathione-dependent lignin degradation pathway found in *Sphingobium* sp. SYK-6 is one example for mediator-independent lignin degradation and is described in chapter 1.8.

LiPs, MnPs, and laccases are extracellular enzymes identified in white-rot fungi and are due to their size unable to penetrate into the plant cell wall to degrade lignin directly.[41] Therefore, these enzymes use small molecules, mediators, to oxidize lignin. LiPs and laccases are also found in some bacterial species, while is this controversial in case of MnPs.[42,43] All three types of enzymes are metal-dependent enzymes using radical reaction mechanisms. Therefore, these enzymes degrade lignin unselectively and it is not possible to produce target molecules out of lignin selectively using them as catalyst. LiPs and MnPs carry a heme prosthetic group, which oxidizes the mediator with the help of hydrogen peroxide. Laccases are copper-containing enzymes with four copper ions in the active centre and use oxygen as oxidative agent.[41] While LiPs and laccases use small organic compounds as redox mediators, MnPs use complexed manganese ions.[41] As mediators for LiPs veratryl alcohol (VA) and 2-chloro-1,4-dimethoxybenzene are known.[44,45] Laccases use different natural mediators, but also accept unnatural ones like TEMPO (2,2,6,6-tetramethyl-1-piperidinyloxy), which is the most effective mediator.[46] Beside the flexibility of laccases in the oxidation of different mediators, also the synthetic oxidation of small organic compounds is known.[47]

After oxidation of the mediator by lignin-degrading enzymes, the mediator diffuses into the plant cell wall and oxidizes lignin there. Due to this oxidation lignin radicals are formed, which are stabilized by various subsequent reactions. These subsequent reactions lead to lignin depolymerization.[41] Figure 1.7 gives an overview on the mechanisms of LiP-, MnP-, and laccase-catalyzed lignin oxidation using mediators.

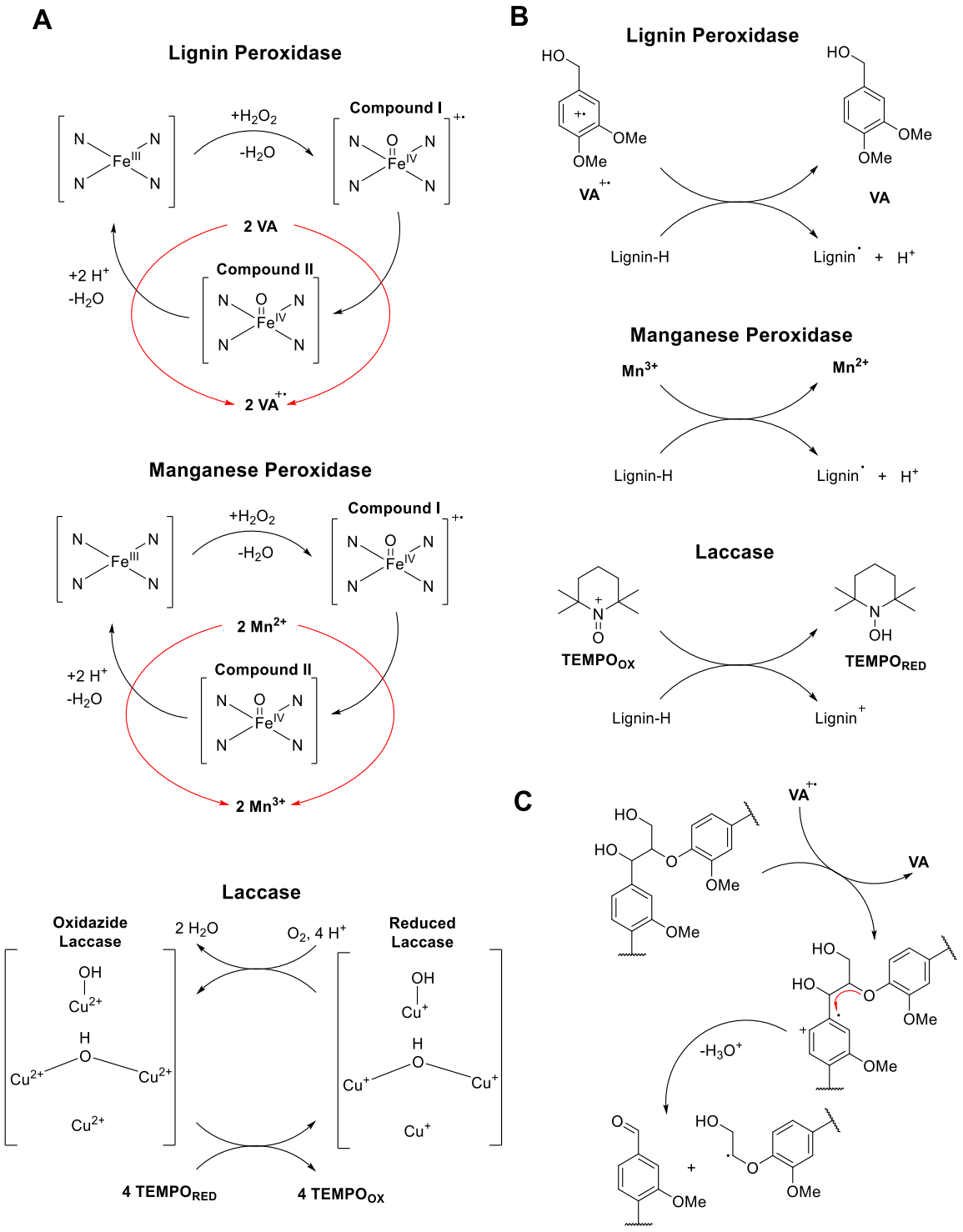


Figure 1.7 Mechanism of mediator-dependent lignin-degrading enzymes. **A** Oxidation of the used mediator by lignin peroxidases, manganese peroxidases, and laccases. **B** Subsequent mediator-dependent oxidation of lignin. The cleavage of a β-4-O aryl ether bond mediated by veratryl alcohol (VA) is described in **C**. Adapted from Pollegioni et al. and Fabbrini et al.[41,46]

1.8 Glutathione-dependent lignin degradation

Glutathione-dependent lignin-degrading enzymes belong to the glutathione S-transferase (GST) superfamily (EC 2.51.18).

1.8.1 Glutathione S-transferases

In general, GSTs are present in all areas of life and contribute up to 1% of the cellular enzymes.[48] GSTs are dimers (often homo-dimers) with one active site per subunit. The subunits consist of a N-terminal thioredoxin domain, formed by β-sheets and α-helices with βαβαββα topology, and a C-terminal helical domain.[48] The enzymes are often involved in the metabolism of xenobiotics such as drugs. This xenobiotic metabolism (Figure 1.8) is divided into three phases: activation and functionalization of the xenobiotic (phase 1), conjugation with another molecule such as glutathione (GSH) to increase the solubility of the xenobiotic (phase 2), and further modification as well as excretion (phase 3).[49]

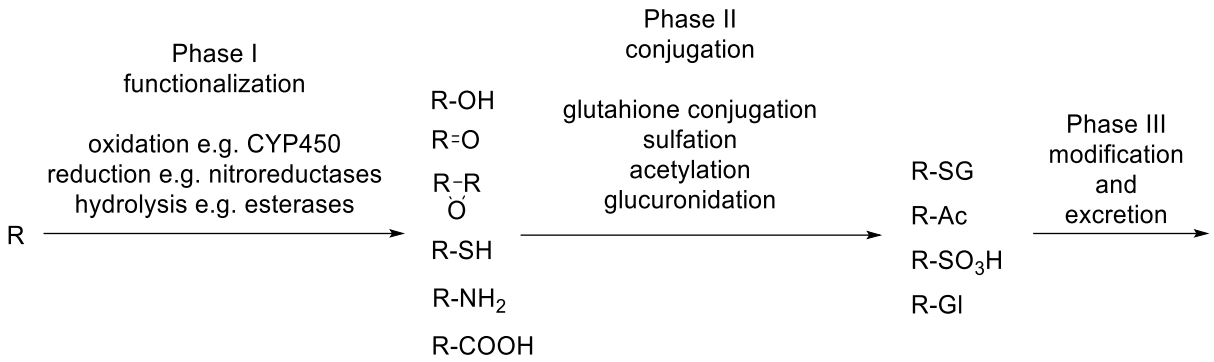


Figure 1.8 General procedure of xenobiotic metabolism.

1.8.2 GSH-dependent lignin degradation pathway

The GSH-dependent lignin degradation enzymes and the corresponding pathway was discovered in *Sphingobium* sp. SYK-6 and was elucidated by Masai and co-workers (Figure 1.9).[50–56] The first step in this enzymatic cascade is the oxidation of the benzylic alcohol group in α-position (see Figure 1.2) to a keto group by the four stereoselective NAD^+-dependent alcohol dehydrogenases (Cα-dehydrogenases) LigD, LigL, LigN, and LigO.[53,57] This keto group is essential for the subsequent ether bond cleavage due to the intramolecular activation of the ether bond and the altered geometry as result of the oxidation. The β-*O*-4 aryl ether cleavage is catalyzed by enantioselective GSH-dependent β-etherases following an S_N2-type mechanism.[55,58] In *Sphingobium* sp. SYK-6, three different β-etherases are known. LigE and LigP cleave ether bonds in (*R*)-configured substrates, while LigF converts the corresponding (*S*)-enantiomers.[58] The product of the ether bond cleavage is a chiral GSH-adduct with inverted stereo-configuration. The thioether bond of this adduct is cleaved by GSH-dependent glutathione lyases, resulting in the formation of oxidized glutathione (GSSG), which is recycled

by the host cell machinery.[55] The glutathione lyase LigG of *Sphingobium* sp. SYK-6 is selective for the (*R*)-thioether,[55] but also converts the (*S*)-configured one with much lower efficiency.[59]

Kontur et al. discovered new groups of β-etherases and glutathione lyases present in some lignin-degrading bacteria species such as *Sphingobium* sp. SYK-6. The novel class of β-etherases are hetero-dimers, in contrast to the classical LigE- and LigF-type β-etherases, while the new group of glutathione lyases, Nu-class glutathione lyases, converts (*S*)-thioether faster than the (*R*)-substrate.[60,61]

Final products of this intracellular pathway, beside GSSG, are small aromatic compounds, which are further metabolized by oxidative ring opening, and are finally introduced into the central carbon metabolism of the host cell for energy production.[56]

Since all enzymes of the GSH-dependent lignin degradation pathway are intracellular enzymes, it is likely that their substrates are predominantly oligomeric lignin degradation products rather than the lignin polymer.[62] Nevertheless, an activity of the enzymes towards lignin polymer has been demonstrated in several cases.[34,63–65]

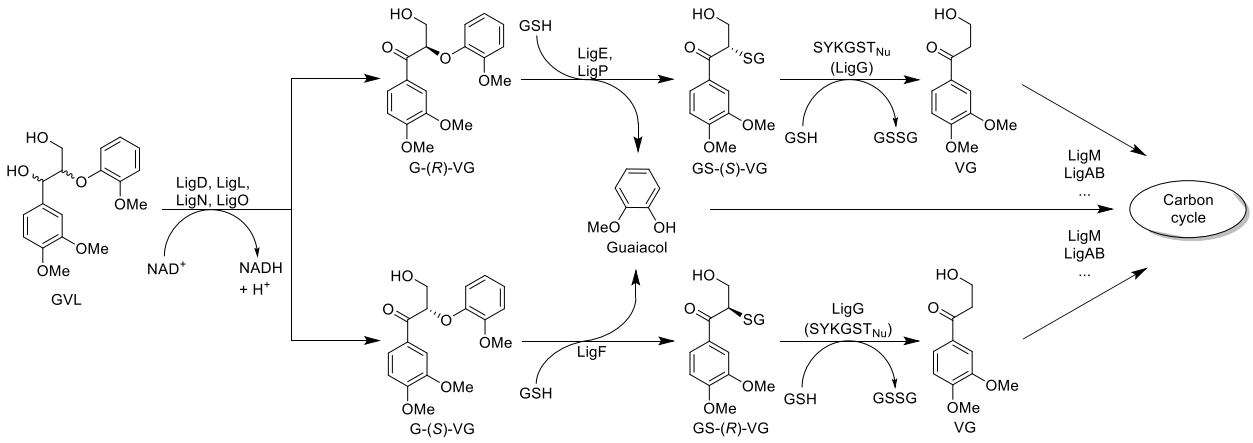


Figure 1.9 GSH-dependent pathway of *Sphingobium* sp. SYK-6 for degradation of β-*O*-4 aryl ether compounds [GVL: β-guaiacyl-α-veratrylglycerol, GVG: β-guaiacyl-α-veratrylglycerone, GS-VG: β-glutathionyl-α-veratrylglycerone, VG: β-deoxy-α-veratrylglycerone, GSH: reduced glutathione, GSSG: oxidized glutathione] Adapted from Husarcíková et al.[62]

Also, fungal enzymes are known to cleave β-*O*-4 aryl ether bonds. In 2018, Marinović et al. identified a GSH-dependent β-etherase in the white-rot fungus *Dichomitus squalens*.[66] Moreover, Otsuka et al. reported in 2003 the discovery of a β-*O*-4 aryl ether cleaving fungal enzyme that is independent of GSH. In contrast to β-etherases, this enzyme is extracellularly located and catalyses a hydrolytic ether bond cleavage. Therefore, this enzyme belongs to a different enzyme class than the bacterial GSH-dependent β-etherases.[67]

1.8.3 β-Etherases

Bacterial β-etherases are much better studied than the two fungal enzymes mentioned above. Nevertheless, still only a few have been characterized, all originating from similar host organisms. Beside the enzymes of *Sphingobium* sp. SYK-6, only bacterial β-etherases of *Novosphingobium* sp. MBES04, *Novosphingobium* sp. PP1Y and *N. aromaticivorans* DSM 12444 have been studied.[30,31] In total, eight (*R*)-selective and six (*S*)-selective β-etherases are known (Table 1.1).[61] The (*R*)-selective enzymes are subdivided in two groups, five LigE-type enzymes and three hetero-dimeric β-etherases. The (*S*)-selective β-etherases consist of five LigF-type β-etherases and Na-LigF-2, which is phylogenetic clearly separated from the LigF-type enzymes.[61] The sequence identities vary in LigE-type β-etherases between 56% and 85%. In (*S*)-selective β-etherases sequence identities among LigF homologs vary between 56% and 96%, while NaLigF-2 is 36% to 42% identical to the other LigF enzymes, demonstrating the clear separation of NaLigF-2 from the other (*S*)-selective enzymes. Within the group of hetero-dimeric β-etherases the sequence identities of the monomers vary between 52% and 74%.

LigE- and LigF-type enzymes display complementary stereoselectivity,[58] but are otherwise biochemically rather equal: All exhibit pH optima in the alkaline range, most of them around pH 9, which is near the pK_a of the GSH thiol group (pK_a 9.65).[30,68] Nevertheless, β-etherases seem to be able to also activate GSH by deprotonation, since some of them are still active at pH 5.[30] The temperature optima of β-etherases are comparably low, usually between 20 °C to 30 °C. Nevertheless, some β-etherases are still active up to 60 °C.[30] Also, solvent stability is reported for β-etherases with tolerating DMSO of at least 25%.[34] Previous studies identified LigE from *Sphingobium* sp. SYK-6 and LigF-NA from *N. aromaticivorans* DSM 12444 as the most active β-etherases among the studied ones.[30] A summary of all currently known β-etherases together with their biochemical data is given in Table 1.1.

Table 1.1 Respective host organisms and biochemical characteristics of known β-etherases.

Enzyme	Origin	GenBank accession number	Activity to GVG (mU·mg^{-1})	pH optimum	Temperature optimum	Selectivity at β-carbon	Reference
LigE	*Sphingobium sp.* SYK-6	WP_014075192	2240	9.0-9.5	30 °C	(*R*)	[30]
LigE-NS	*Novosphingobium* sp. PP1Y	WP_013832481	140	9.0-9.5	<20 °C	(*R*)	[30]
LigE-NA	*N. aromaticivorans* DSM 12444	WP_011446047	150	9.0-9.5	<20 °C	(*R*)	[30]
GST5	*Novosphingobium* sp. MBES04	WP_039391125	170	7.0-8.0	40 °C	(*R*)	[68]
LigF	*Sphingobium sp.* SYK-6	WP_014075191	530	9.0-9.5	25 °C	(*S*)	[30]
LigF-NS	*Novosphingobium* sp. PP1Y	WP_013832480	300	9.0-9.5	<20 °C	(*S*)	[30]
LigF-NA	*N. aromaticivorans* DSM 12444	WP_041551020	2920	9.0-9.5	25 °C	(*S*)	[30]
NaLigF1[a]	*N. aromaticivorans* DSM 12444	ABD26530	n.r.	n.r.	n.r.	(*S*)	[31]
NaLigF2	*N. aromaticivorans* DSM 12444	ABD27301	n.r.	n.r.	n.r.	(*S*)	[31]
GST4	*Novosphingobium* sp. MBES04	WP_039391123	370	9	35 °C	(*S*)	[68]
LigP	*Sphingobium* sp. SYK-6	WP_014077574	6.2	9.0-9.5	25 °C	(*R*)	[30]
BaeAB	*N. aromaticivorans* 12444Δ1879	BaeA WP_011446513 BaeB WP_011446512	(comparable to LigE-NA)	n.r.	n.r.	(*R*)	[61]
PP1Y-BaeAB	*Novosphingobium* sp. PP1Y	BaeA WP_013832467 BaeB WP_013832468	n.r.	n.r.	n.r.	(*R*)	[61]
Sxe-BaeAB	*S. xenophagum* NBRC 107872	BaeA WP_019052344 BaeB WP_019052345	n.r.	n.r.	n.r.	(*R*)	[61]
Ds-GST1	*D. squalens* LYAD-421 SS1	XP_007363869	0.11	n.r.	n.r.	(*S*)	[66]

[a] Enzyme identical to LigF-NA except for 11 additional amino acids at the N-terminus

n.r.: not reported

In terms of structure, LigE- and LigF-type β-etherases are homodimers with a molecular weight of around 60 kDa (for the dimer). Helmich et al. published in 2016 the crystal structures of LigE and LigF.[69] Like many other enzymes of the GST superfamily, LigE and LigF also consist of a N-terminal thioredoxin domain and a C-terminal helical domain which are connected by a short linker.[69] The thioredoxin domain is responsible for GSH binding. Amino acid residues of both domains form the active site of LigE and LigF.[69] The active site in LigE is located in a surface-exposed cleft, whereas in LigF the active site is placed in a tunnel-like structure (Figure 1.10). In LigE only residues of the helical domain form the dimer interface, whereas in LigF amino acids of both domains contribute to dimer formation. This is also the reason why LigE and LigF belong to different GST classes. LigE is member of the GSTFuA class. Based on phylogenetics, LigF could be also grouped into this class, but the dimer interface of LigF is different from typical GSTFuA class members. So, LigF is suggested to belong to a new class of GSTs.[69] The exact substrate binding in β-etherases is not yet understood, since no crystal structure with bound substrate could be obtained. However, structures with co-crystallized GSH are available.[69]

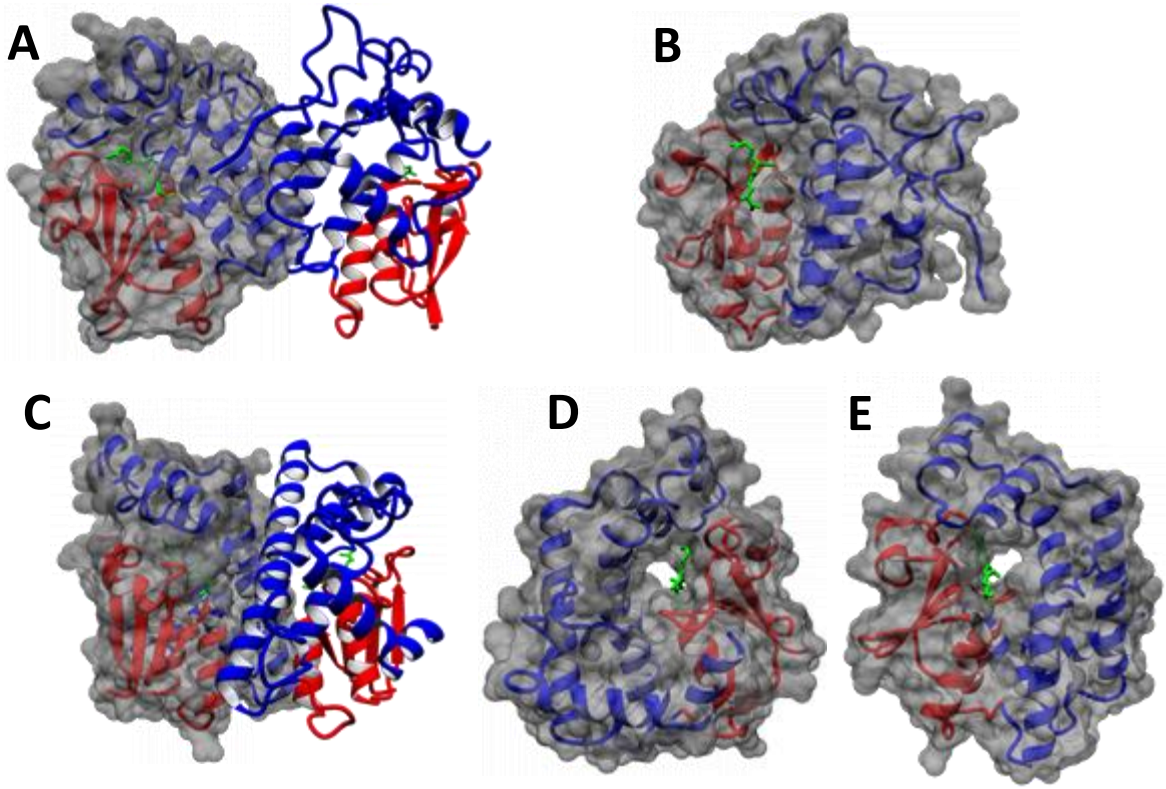


Figure 1.10 Crystal structures of LigE (PDB: 4YAN; **A** dimer**, B** monomer) and LigF (PDB: 4XT0; **C** dimer, **D** monomer with tunnel entrance. **E** opposite monomer side towards the dimer interface). The co-crystallized cofactor GSH is visualized as green sticks. The thioredoxin domain is shown in red, while the helical domain is in blue.

In 2019, Kontur et al. discovered in *N. aromaticivorans* a hetero-dimeric enzyme with β-etherases activity, named BaeAB.[61] Both polypeptide chains can be expressed separately in a cell-free expression system, however, the separate monomers were shown to be inactive. If the two monomers are mixed, the activity is restored. Separate expression of both monomers in *E. coli* did not yield active enzyme, whereas the combined expression of BaeAB results in active enzyme.[61] BaeAB is phylogenetically more related to LigF-type enzymes, but displays the same stereoselectivity as LigE-type enzymes.[61] BaeA is sequence wise 18% and 35% identical to LigE and LigF, respectively, whereas BaeB is 17% and 25% identical to LigE and LigF, respectively. The absolute activity of BaeAB was not determined, but is comparable to the activity of LigE-NA. Further hetero-dimeric β-etherases were also identified in *Novosphingobium* sp. PP1Y and *S. xenophagum*.[61]

1.8.4 Glutathione lyases

In the third step of the GSH-dependent lignin degradation pathway glutathione lyases catalyse the cleavage of the thioether linkage of the GSH-adduct with the release of GSSG. Bacterial glutathione lyases have been identified in all organisms with characterized β-etherases.[59,60] In 2015, Picart et al. characterized additionally a glutathione lyase from *T. denitrificans* ATCC 25259 with activity on GSH-adducts formed by β-etherase catalysis.[59] This bacterium, however, is not known for lignin degradation, indicating that lignin degradation products are not the natural substrates of this enzymes.[62]

Analogous to β-etherases, the glutathione lyases of the β-*O*-4 aryl ether degradation pathway are also subdivided in two classes: the Omega-class and a recently identified Nu-class glutathione lyases.[60] Omega-class enzymes are (*R*)-selective, whereas the enzymes of the Nu-class preferentially converts the (*S*)-enantiomer of the GSH-adduct.[59,60] Both enzyme types, however, are also able to convert the respective other substrate enantiomer. Hence, glutathione lyases are less enantioselective than β-etherases.[59] The pH and temperature optima of glutathione lyases as well as their stability against heat and organic solvents is comparable to β-etherases, with alkaline pH optima and low optimal reaction temperature. Nevertheless, some enzymes such as LigG-TD display also higher thermostability.[59,63,68] The activity of glutathione lyases is difficult to compare based on literature data since the different authors did not use the same substrates for activity determination.[59,60,63,68] Comparing the activity of LigG, LigG-NS and LigG-TD, Picart et al. showed that LigG is the most active one and

LigG-TD the least stereoselective one of these three glutathione lyases.[59] Table 1.2 gives an overview of known glutathione lyases.

Table 1.2: Host organisms and biochemical characteristics of known glutathione lyases.

Enzyme	Origin	GenBank accession number	pH optimum	Temp. optimum	Selectivity (β carbon)	Reference
LigG	*Sphingobium* sp. SYK-6	BAA77216	9.0	20 °C	(*R*)-preference	[59]
LigG-NS	*Novosphingobium* sp. PP1Y	CCA92089	9.0	20 °C	(*R*)-preference	[59]
LigG-TD	*T. denitrificans* ATCC 25259	WP_011311562	9.5	20 °C	(*R*)-preference	[59]
GST6	*Novosphingobium* sp. MBES04	GAM05532	n.r.	n.r.	(*R*)-preference	[68]
GST3	*Novosphingobium* sp. MBES04	WP_039391121	8.0	25-30 °C	(*S*)-preference	[63,68]
$NaGST_{Nu}$	*N. aromaticivorans* DSM 12444	WP_011446237	n.r.	n.r.	(*S*)-preference	[60]
$SYKGST_{Nu}$	*Sphingobium* sp. SYK-6	BAK65087	n.r.	n.r.	(*S*)-preference	[60]

n.r.: not reported

As already mentioned, glutathione lyases are also members of different GST subgroups. The LigG homologs belong to the Omega-class of GSTs,[70] while Nu-class glutathione lyases such as $NaGST_{Nu}$ are members of the Nu-class of GSTs.[60] The dimer interface of LigG and $NaGST_{Nu}$, the two glutathione lyases with known crystal structure, consists of regions of both domains, the N-terminal thioredoxin domain and the C-terminal helical domain. The dimer interface of $NaGST_{Nu}$, however, is much larger than the respective interface of LigG.[60,70]

The active site of Omega-class enzymes is found in a surface exposed cleft as in the case of β-etherase LigE type, whereas in $NaGST_{Nu}$ the active site is located in a tunnel structure as found for β-etherase LigF.[60,71] This similarity is very interesting, since it correlates with the enzymes' stereospecificity on the β-carbon of the respective substrate: LigE and LigG are (*R*)-selective, whereas LigF and $NaGST_{Nu}$ are (*S*)-selective.

In the case of LigG, there is some knowledge available about the exact substrate binding, since Pereira et al. have been able to co-crystallize LigG with the substrate analogue β-glutathionyl-acetoveratrone (GS-AV).[71] In this structure, GS-AV shows pi-pi interaction with the aromatic side chain of Y113. Mutagenesis this of residue led to a nearly complete loss of activity. The thioether bond of the substrate is located near C15. Hence this cysteine could perform a

nucleophilic attack on the thioether substrate, resulting in thioether bond cleavage and formation of a disulfide bridge between C15 and GSH. Based on this, Pereira et al. postulated a two-step mechanism for LigG. In the second step, a second GSH molecule would cleave the disulfide bridge of the GSH-enzyme-adduct with release of GSSG.[71] Beside this a second crystal structure of LigG with GSH-bound, published by Meux et al., is known.[70] In both crystal structure, free GSH and the GSH-part of the substrate GS-AV are bound in the same way to the thioredoxin domain (Figure 1.11). In the LigG structure with co-crystallized GSH, C15 forms a disulfide bridge with the free GSH.

$NaGST_{Nu}$ could be crystallized together with two bound GSH molecules, either in oxidized or reduced form. Based on modelling results, Kontur et al. postulated, that GSH and substrate would bind to $NaGST_{Nu}$ at the same time. The enzyme activates then the thiol group of GSH for nucleophilic attack on the thioether bond of the substrate resulting in thioether, which cleavage and a release of GSSG.[60] A summary of both postulated mechanisms for glutathione lyases is shown in Figure 3.15.

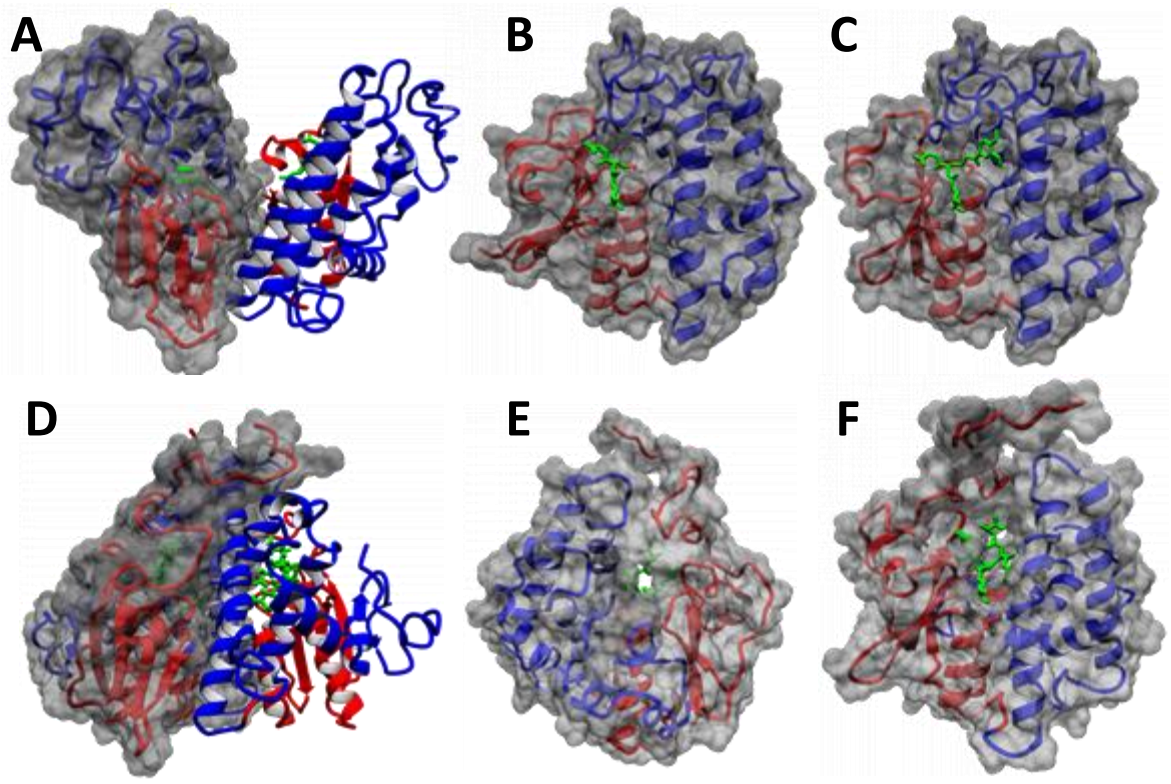


Figure 1.11 Crystal structures of glutathione lyase LigG co-crystallized with GSH (PDB: 4G10; **A** dimer, **B** monomer) and co-crystallized with substrate GS-AV (PDB: 4YAV; **C** monomer), as well as crystal structure of $NaGST_{Nu}$ co-crystallized with two GSH molecules per active site (PDB: 5UUO; **D** dimer, **E** monomer with tunnel entrance. **F** opposite monomer side towards the dimer interface). The bound molecules GSH and GS-AV are colored in green. The thioredoxin domain is shown in red, while the helical domain is in blue.

1.8.5 Applications of GSH-dependent lignin-degrading enzymes

β-Etherases and glutathione lyases were used in different enzyme cascades for the degradation of lignin polymer. For oxidation of the α-hydroxyl groups in lignin, which is necessary for β-etherase activity, Cα-dehydrogenases of the GSH-dependent lignin degradation pathway or a laccase were used.[34,65] The first system was described by Reiter et al. consisting of LigD, LigF, and LigG. For internal cofactor regeneration a NADH-dependent glutathione reductase AvGR from *Allochromatium vinosum* was added resulting in a self-sufficient system (Figure 1.12).[65] The process yielded only low amounts of mono-aromatic products, based on the usage of enzymes specific for one enantiomer in α- and β-position only. Therefore, just the (α*R*,β*S*)-configurated aryl ether bond could be cleaved by the used system, 25% of the total β-4-O-aryl ether bond in lignin.[62]

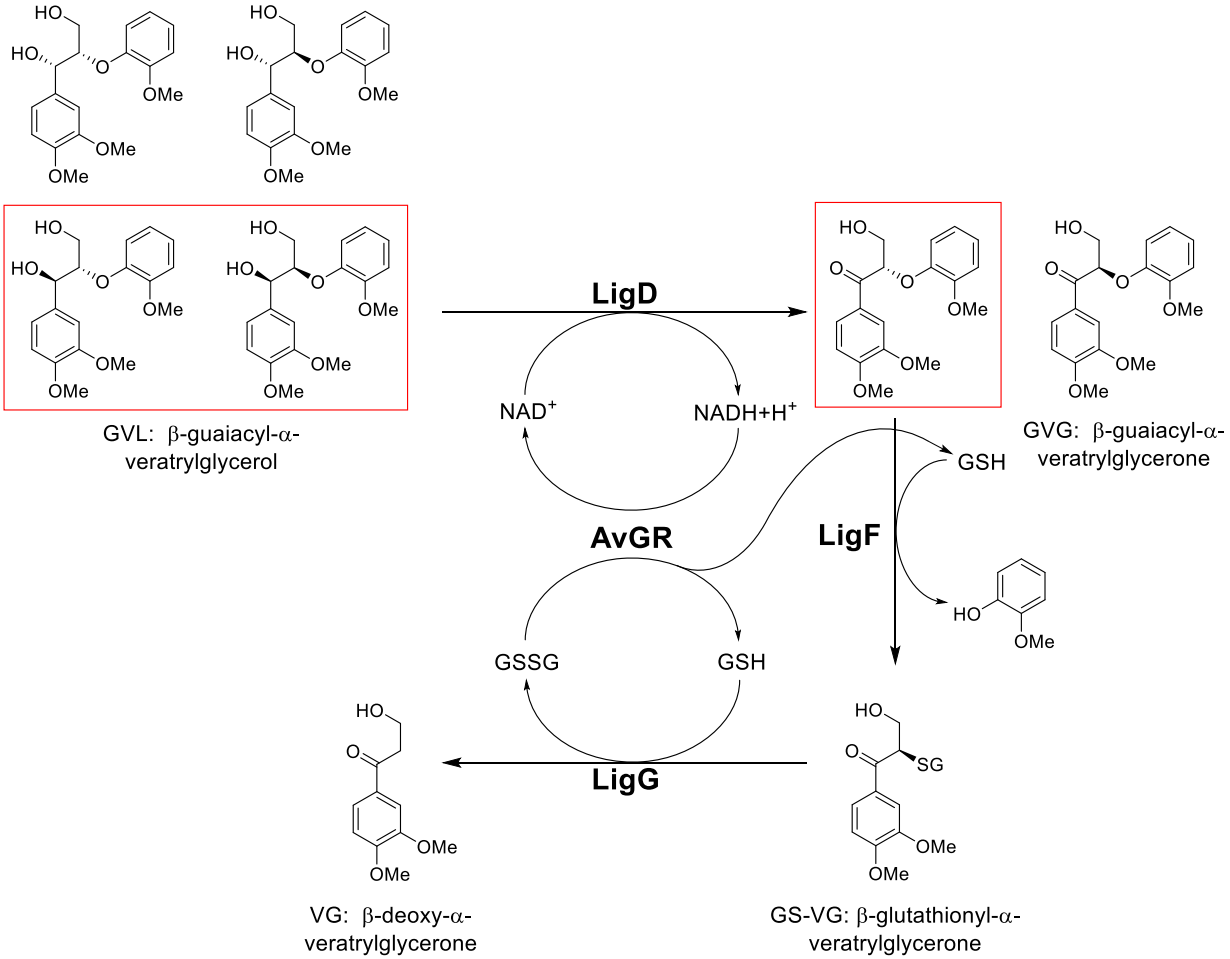


Figure 1.12 Reaction scheme of the self-sufficient lignin-degrading process reported by Reiter et al. shown with GVL as starting compound.[65] In this process, LigD oxidizes GVL, which is subsequently cleaved by LigF and LigG. The glutathione lyase AvGR leads to a self-sufficient cofactor regeneration of NAD^+ and GSH. Due to the stereospecificity of LigD and LigF only (α*R*)-GVL and (β*S*)-GVG are converted in this process (red box). Adapted from Reiter et al.[65]

Ohta et al. and Gall et al. developed this lignin degradation system further by combining two enantio-complementary Cα-dehydrogenases and two enantio-complementary β-etherases with an unselective glutathione lyase.[32,63] The system of Ohta et al. consisted of enzymes found in *Novosphingobium* sp. MBES04 (Cα-dehydrogenases: SDSR3, SDSR5; β-etherases: GST4, GST5; glutathione lyase: GST3), whereas Gall et al. combined enzymes of *Sphingobium* sp. SYK-6 (Cα-dehydrogenases: LigD, LigN; β-etherases: LigE, LigF), *N. aromaticivorans* (glutathione lyase: $NaGST_{Nu}$), and the glutathione reductase AvGR.[32,63]

Beside these three one-pot cascade systems, Picart et al. designed a two-pot, two-step system for lignin degradation. In the first step lignin is oxidized using a laccase mutant from *T. versicolor*, Lcc3 M3, and the mediator violuric acid. Afterwards, the β-etherases LigE and LigF-NA together with the glutathione lyase LigG-TD were used for the cleavage of the aryl ether bond.[34]

The yield of mono-aromatic products obtained in these enzymatic lignin depolymerizations is generally increased by using enzymes with complementary enantioselectivity. While the system by Reiter et al. yielded less than 2 wt% of small aromatics, the Ohta system produced 6.6 wt%, and Gall et al. as well as Picart et al. 12.5 wt%. Figure 1.13 summarizes the different reported GSH-dependent lignin degradation processes.[32,34,63,65]

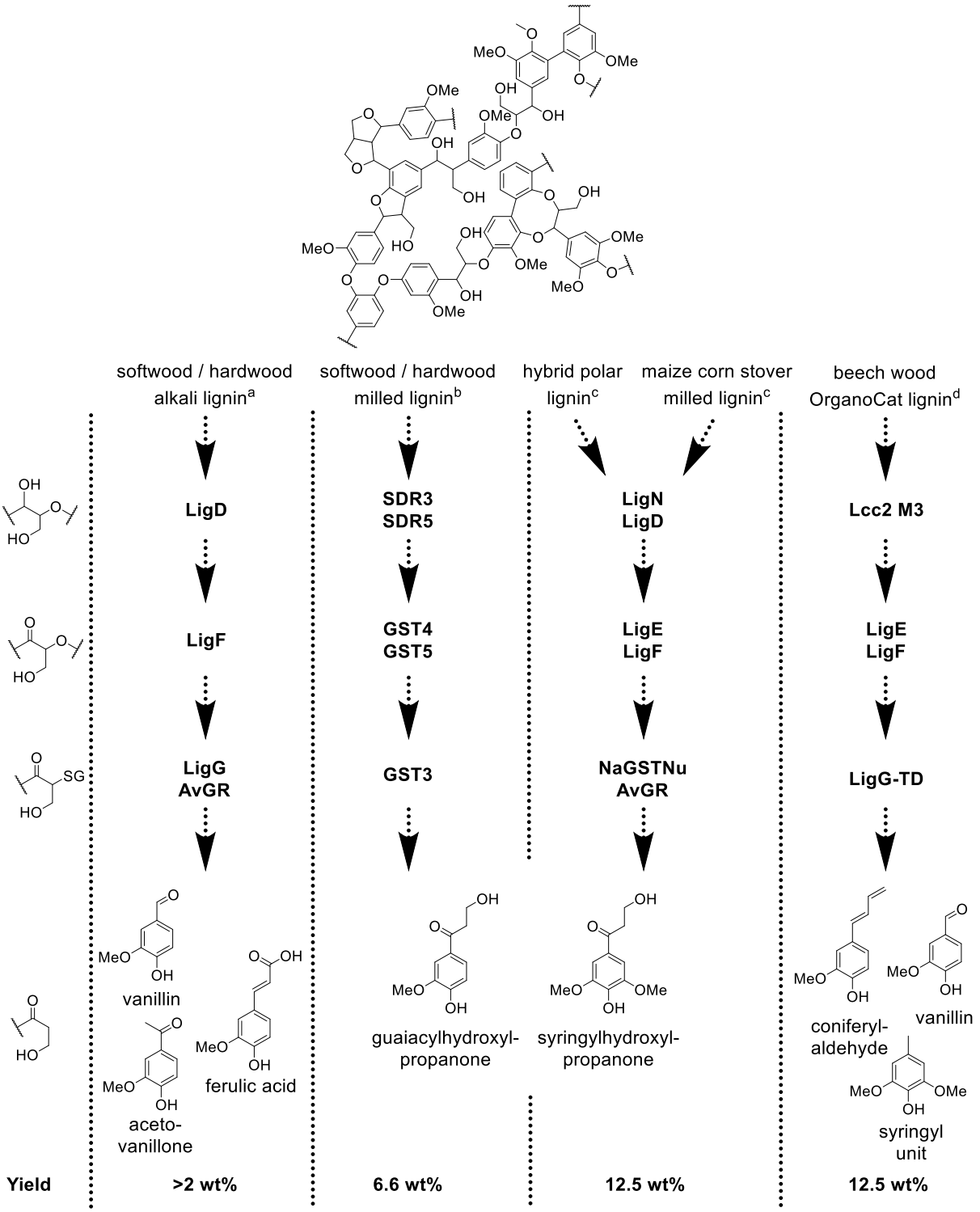


Figure 1.13 Summarize of GSH-dependent lignin degradation processes. Shown are the used enzymes, main products and yields of small aromatic products. [[a]: Reiter et al.[65], [b]: Ohta et al.[63], [c]: Gall et al.[32], [d]: Picart et al.[34]].

1.9 Peptide pattern recognition

The search for new, uncharacterized enzymes with desired function is an important part of modern biosciences. Aim of this line of research is usually to find enzymes with better or new properties than currently known enzymes for a desired application. One source of these enzymes of interest are public sequence databases, listing millions of uncharacterized enzymes, e.g. by sequencing environmental samples. Often used for public database mining is the BLAST tool (*Basic Local Alignment Search Tool*).[72] BLAST is an online sever, which searches in public databases based on one or more protein sequences (in case of BLASTP), given by the user as query. The BLAST algorithm produces very short sequence fragments out of the query sequence, which are used for mining the databases. Based on these short sequence alignments are formed, which are expanded to maximum length. The result of this process is statistically ranked.[72] Hence, a BLAST search results in the identification of sequence-wise related proteins to the query sequence.

In the case of β-etherases and glutathione lyases as members of the GST superfamily, a BLAST search typically yields thousands of sequences, all encoding GSTs, due to the distribution of this enzyme family.[48] Not all of them, however, encode for β-etherases or glutathione lyases. Picart et al. tested for example the enzyme LigP-SC from *Sorangium cellulosum* So Ce56, which is homologous to LigP but does not possess β-etherases activity.[30] Likewise, RpHypGST from *Rhodopseudomonas palustris*, tested by Gall et al., did also not display β-etherase activity.[31] Specifically, if more distantly related homologues to the query sequence are desired to increase the probability of identifying enzymes with interesting new properties, the risk is significantly increased that those sequences do not anymore encode an enzyme with desired function.

A relatively new approach to identify homologs is the peptide pattern recognition (PPR) tool, introduced by Busk and Lange.[73] PPR is a non-alignment-based algorithm, which groups protein sequence databases based on the conservation of small peptides. The conservation of small peptides in a group of proteins, especially around the active side, can be an indication for sharing the same function, even if the proteins are rather unrelated in terms of their sequence.[73]

The PPR requires a database of protein sequences, for example generated by BLASTP search in public databases, as starting point. The algorithm lists all possible peptides with a defined length (e.g. 6 AA) and checks that of these peptides are present in the database more than

once. PPR then groups protein sequences together which share more small peptides than the defined cut-off value. For the successful use of PPR it is required to know a set of proteins (3 to 5) with the desired function. If the algorithm groups all these proteins in the same group, it is likely that also the other proteins in the group share the same function.[73]

Busk and Lange tested the PPR algorithm by predicting the functions of glycoside hydrolases. The function of 118 enzymes of class GH5 and 540 enzymes of class GH13 was predicted with 97% and 82% accuracy, respectively.[73] Also, a new the β-glucosidase was identified in *Mucor circinelloides* with the help of PPR.[74] Moreover, the PPR algorithm was successfully applied to other enzyme groups active on sugar molecules such as polysaccharide monooxygenases, and glucuronoyl esterases.[75,76]

1.10 Protein engineering

1.10.1 Biocatalysis in nature and industry

Enzymes are interesting biocatalysts for chemical reactions. Enzymes are optimized by nature for chemical reactions occurring in living cells or other natural systems. In most cases, this means aqueous environment, low substrate concentrations (µM to mM), neutral pH and ambient temperature.[77,78] In many applications in the lab or in industry the conditions, under which enzymes are used, differs from these natural conditions.

In contrast, chemo-catalysts usually work in organic solvents and at high substrate concentrations.[79] Nevertheless, enzymes are interesting since they are usually more selective (regio- or enantioselective) than chemo-catalysts and in some cases catalyse reactions under conditions which are impossible using chemo-catalysts.[80] One example for such a reaction is the deacylation of Penicillin G in the synthesis of semi-synthetic penicillin derivatives. The selective hydrolysis of the amide bond in Penicillin G to yield 6-aminopenicillic acid catalyzed by Penicillin acylase runs in water and at room temperature (or higher), whereas the chemical deacylation requires hazardous chemicals, solvents and very low temperatures (Figure 1.14). This leads to high amounts of waste and high energy costs in the chemical process.[81]

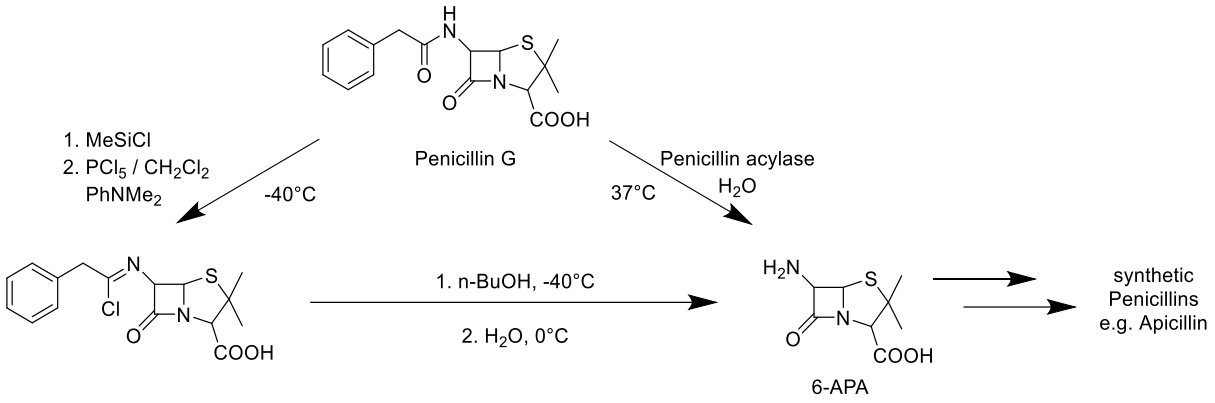


Figure 1.14 Reaction scheme of the enzymatic and chemical deacylation of Penicillin G in the synthesis of semi-synthetic penicillin derivatives. Adapted from Sheldon et al.[81]

The main challenges in the use of enzymes as catalysts for chemical reactions are their rather low stabilities, their often low substrate concentration tolerance and the lack of known enzymes for desired chemical reactions. Therefore, natural enzymes usually need to be adapted for practical applications. Stability, selectivity, activity, substrate concentration tolerance and solvent tolerance are the main parameters of interest in protein engineering campaigns.[80]

To change the amino acid composition of a protein and therefore also the properties of this protein the DNA sequence of the protein-encoding gene is mutated. For this, two different strategies are known in protein engineering: random mutagenesis or rational design.

1.10.2 Random mutagenesis and directed evolution

In random mutagenesis, mutations are introduced randomly in the whole gene or a gene fragment of interest. The resulting mutant library is then screened for a desired property and only improved variants are sequenced afterwards. In directed evolution, this process is performed over several rounds, taking the best variant of the previous mutagenesis round as template for the next (Figure 1.15). By this way only a few mutations are introduced per round and the enzyme of interest can be significantly improved in the desired property.[80] Frances Arnold (Nobel price 2018 for chemistry) and co-workers demonstrated first the power of directed evolution by introducing 10 mutations in a serine protease within three rounds of random mutagenesis, which resulted in a 256 times more active variant.[82]

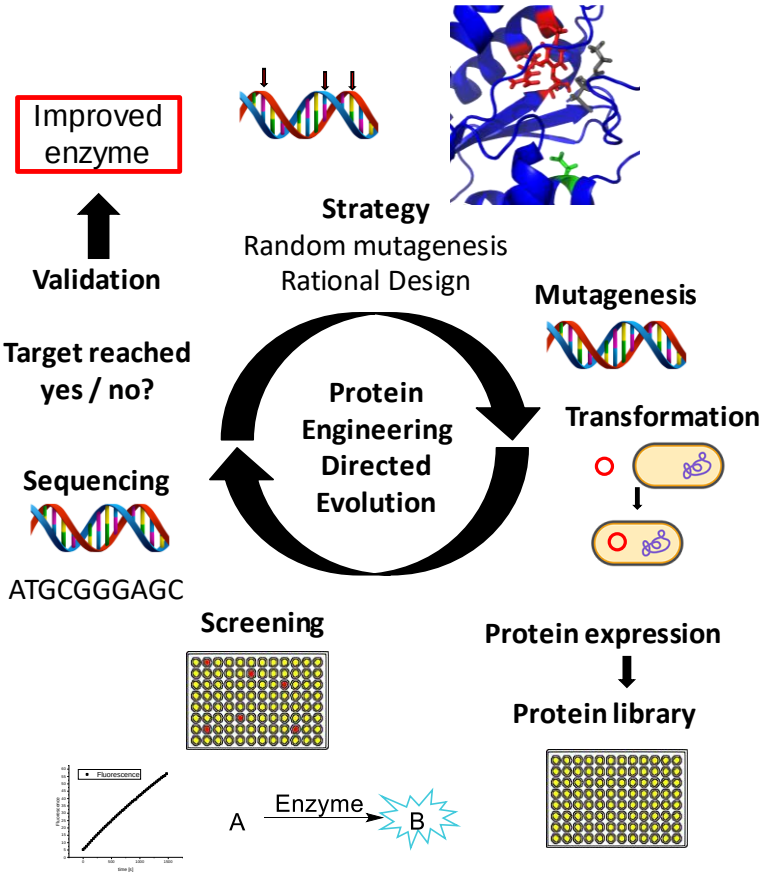


Figure 1.15 Workflow of directed evolution.

Methods such as error-prone PCR or gene shuffling are distinguished in random mutagenesis (Figure 1.16). In error-prone PCR standard fidelity DNA polymerase (usually Taq-polymerase) is used under conditions, which are promoting higher error frequencies during the gene amplification mutations (e.g. addition of manganese ions, unbalanced nucleotide concentrations).[83] In DNA shuffling, a set of homologous genes are partially digested with DNaseI into random fragments. Overlapping fragments are then combined to full-length genes by a PCR reaction without primers.[84]

Both techniques introduce random mutations over the entire gene or fragment of interest and do not require any additional information about the enzyme beside the gene sequence.[85] The main drawback of these methods is the required high screening effort to find improved variants. Furthermore, in random mutagenesis campaigns the point where no further improvement can be accepted cannot be predicted, since the possible amino acid combinations are nearly endless. Therefore, both is possible: more screening is required for further improvement or the best possible variant already known. On the other hand, these methods enable the identification of mutants often overlooked in rational approaches, since

rational approaches are often restricted to specific areas of the enzyme such as the active site.[86]

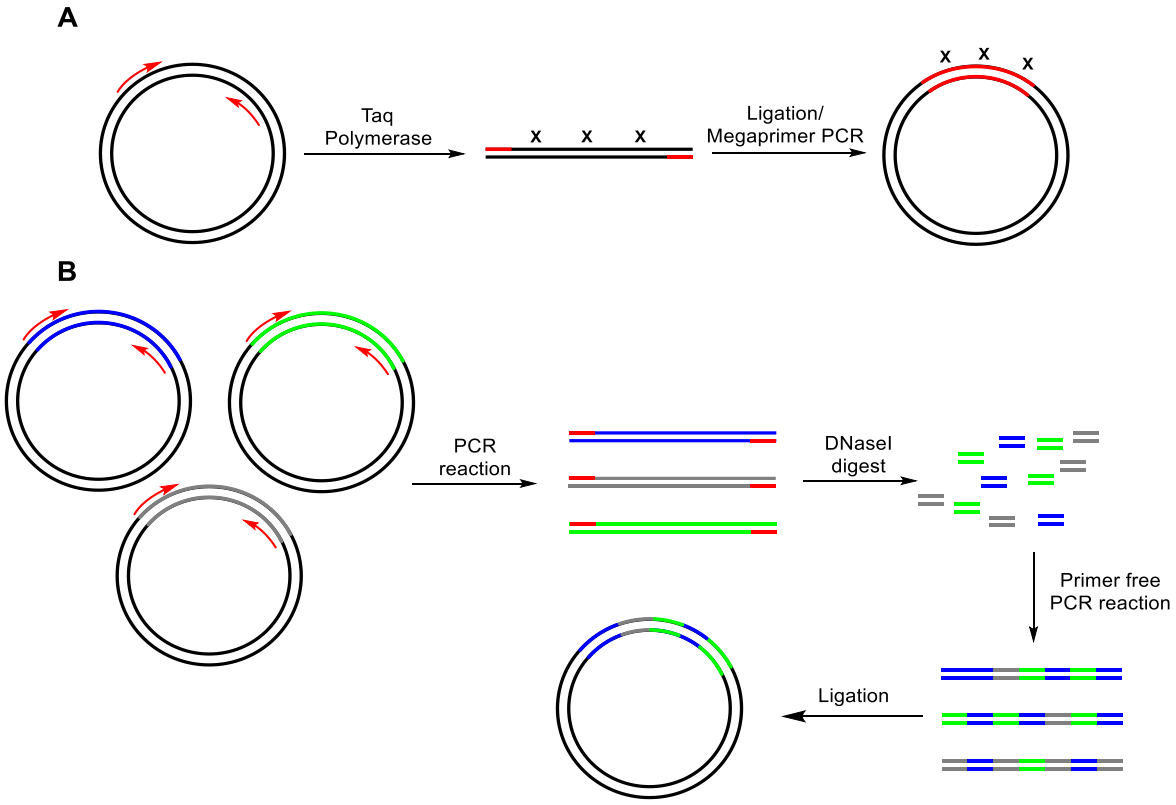


Figure 1.16 Scheme of error prone PCR (**A**) and DNA shuffling (**B**) mutagenesis. Mutations are marked with **X**.

1.10.3 Rational and semi-rational design

In rational protein design, additional information about the enzyme structure and mechanism are used to improve enzyme characteristics with a limited number of predicted mutations. Site-directed mutagenesis by PCR with primers carrying the desired mutations are used to introduce the predicted mutations into the gene of interest. Quikchange or the Q5 protocol are common methods for site directed mutagenesis (Figure 1.17). In the Quikchange protocol, a Pfu DNA polymerase and complementary primers carrying the mutation in the middle are used. The whole vector is amplified in the PCR reaction producing a double nicked PCR product.[87] The Q5 protocol uses non-complementary primer pairs that bind next to each other in the gene. The product of the PCR reaction using the high fidelity Q5-DNA polymerase is linear DNA fragment with blunt ends, which is subsequently cyclized in vitro.[88] The wildtype template is in both approaches removed specifically due to the *E. coli* DNA-methylation pattern by DpnI digest.

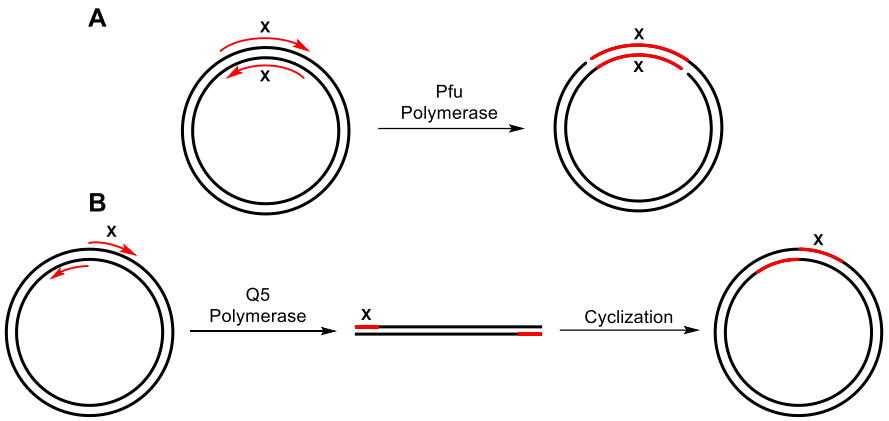


Figure 1.17 Scheme of Quikchange (**A**) and Q5 (**B**) mutagenesis. Mutations are marked with **X**. Adapted from Agilent Technologies (Waldbronn, Germany) and NEB (Frankfurt am Main, Germany).[87,88]

The big challenge in rational design is the prediction of beneficial amino acid exchanges. With the help of structural information and alanine scans, positions in the gene with a great influence on the enzyme property of interest can be identified. However, it is still difficult to predict which amino acid should be incorporated to achieve an improvement. Nowadays, this problem is usually tackled by combination of rational design and random mutagenesis. Site-directed saturation mutagenesis (SSM), a semi rational approach, uses degenerated primer to cover every possible mutation to a desired position in the gene. As result high diversity with moderate screening effort is created. The big advantage of SSM is that the number of clones needed to screen all possible mutation can be calculated. Therefore, the experiments and the end point of the experiment, where no further improvement is expectable, is predictable in contrast to random mutagenesis approaches.[89] The semi-rational analogue to directed evolution is called iterative saturation mutagenesis (ISM). ISM leads often to comparable enzyme improvement like generated with random mutagenesis with highly reduced screening effort.[90]

1.11 Aim of this thesis

The focus of this PhD thesis is the investigation of glutathione-S-transferases featuring the ability to depolymerize lignin, the key step for the utilization of lignin as renewable source for biofuel or chemicals production. In contrast to other technical or enzymatic processes which lead to unselective lignin depolymerization, glutathione-dependent β-etherases and glutathione lyases are able to specifically cleave the β-*O*-4-aryl ether bond in lignin and, therefore, generate a defined product spectrum of small aromatics.

Despite these very interesting enzymatic features, the number of known β-etherases and glutathione lyases is very limited. Therefore, one aim of this project was to increase the number of available β-etherases as well as to identify more diverse members of this enzyme family. To achieve this, different public database mining strategies were used, and a representative set of novel enzymes was characterized. The increased number of β-etherase sequences combined with structural information should be used afterwards to identify conserved amino acids and catalytically important regions within the enzymes.

In the second part of this project, engineering of the glutathione lyase LigG-TD with the aim to improve its activity while reducing its enantioselectivity. Since, no high-throughput screening assay for glutathione lyases was available at the beginning of the project, the first goal was to develop an absorbance- or fluorescence-based assay for the determination of glutathione lyase activity in microtiter-plate format. This assay was then to be applied in the screening of different LigG-TD mutant libraries for the identification of more active variants.

The biochemical characterization of novel β-etherases as well as the activity screening of glutathione lyases requires lignin model compounds as substrate which are not commercially available. Hence, multi-step syntheses towards different model compounds had to be established and optimized as well.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals were of analytical grade or higher quality and purchased from Sigma-Aldrich (München, Germany), Carl Roth (Karlsruhe, Germany), Acros Organics (Geel, Belgium), VWR (Darmstadt, Germany), Thermo Fisher Scientific (Waltham, USA-MA), Alfa Aesar (Karlsruhe, Germany), New England BioLabs (Frankfurt am Main, Germany), Omega Bio-tek (Norcross, USA-GA)

2.1.2 Equipment

Table 2.1 Equipment used in this thesis.

Equipment	Producer
Chromatography systems	
ÄKTA Pure protein purification system	GE Healthcare (Solingen, Germany)
ÄKTA Start protein purification system	GE Healthcare (Solingen, Germany)
Nexera XR HPLC System -LC-20AD XR -LC-20AD XR -Sil-20 AC XR -SPD-M20A -CTO-20 AC -CBM-20A	Shimadzu Deutschland GMBH (Duisburg, DE)
UV-Vis analytic	
CARY 60 Bio UV-Vis	Agilent Technologies (Waldbronn, Germany)
CLARIOstar microtiter plate reader	BMG LABTECH Gmbh (Ortenberg, DE)
NanoPhotometer NP80	Implen (München, Germany)
Centrifuges	
Heraeus Fresco 21 microcentrifuge	Thermo Fisher Scientific (Waltham, USA-MA)
Heraeus Multifuge X3R centrifuge	Thermo Fisher Scientific (Waltham, USA-MA)
Mega Star 3.0 R centrifuge	VWR (Darmstadt, Germany)
Micro Star 17 centrifuge	VWR (Darmstadt, Germany)
Incubators	
INCU-Line Incubator	VWR (Darmstadt, Germany)
MaxQ 8000 shaker	Thermo Fisher Scientific (Waltham, USA-MA)
Minitron Infors HT	Infors AG (Einsbach, Germany)
ThermoMixer C	Eppendorf (Hamburg, Germany)
VWR Incubating microplate shaker	VWR (Darmstadt, Germany)
Electrophoresis equipment	
FastGene blue/green GelPic LED Box	NIPPON Genetics europe (Düren, Germany)
Mini PROTEAN SDS-PAGE system	Bio-Rad Laboratories (Feldkirchen, Germany)
Owl EasyCast B1A Mini Gel-electrophoresis systems	Thermo Fisher Scientific (Waltham, USA-MA)
MS 300V power supply	Major Science (Saratoga, USA-CA)

Equipment	Producer
Scale	
Entris 224i-1S precision balances	Sartorius (Göttingen, Germany)
Entris 3202i-1S precision balances	Sartorius (Göttingen, Germany)
PCR Machines	
CFX96 real time system c1000 thermal cycler	Bio-Rad Laboratories (Feldkirchen, Germany)
PeqSTAR thermocycler	VWR (Darmstadt, Germany)
Vacuum systems	
Rotavapor® R-300	BÜCHI Labortechnik Gmbh (Essen, Germany)
RZ6 high vacuum pump	Vacuubrand (Wertheim, Germany)
Other lab equipment	
Arium® pro ultrapure water system	Sartorius (Göttingen, Germany)
Biological safety cabinet Safe 2020	Thermo Fisher Scientific (Waltham, USA-MA)
Eppendorf Research® plus pipettes	Eppendorf (Hamburg, Germany)
Fisherbrand™Model 120 Sonic dismembrator	Thermo Fisher Scientific (Waltham, USA-MA)
Sonorex Digitec	BANDELIN electronic Gmbh & Co. KG (Berlin, Germany)
Vortex-Genie2	Scientific Industries Inc (Bohemia, USA-NY)
Chemical analytic used in cooperation	
Bruker Avance AVII 600 NMR spectrometer	Bruker Corporation (Rheinstetten, Germany)
Finnigan LCQ Deca ESI mass spectrometer	Thermo Fisher Scientific (Waltham, USA-MA)
LTQ-Orbitrap Velos ESI-HRMS mass spectrometer	Thermo Fisher Scientific (Waltham, USA-MA)
LC MS: -UHPLC system Ultimate3000RS -Kinetex C18 column (1,7µm 100A, 150x2,1mm) -Maxis HD UHR-TOF -Apollo II Elektrospray source	 Thermo Fisher Scientific (Waltham, USA-MA) Phenomenex (Torrance, USA-CS) Bruker Corporation (Rheinstetten, Germany) Bruker Corporation (Rheinstetten, Germany)
Monolith NT.115 micro scale thermophoresis	NanoTemper Technologies GmbH (München, Germany)

2.2 Computer programs and web servers

Table 2.2 List of used computer programs and web servers.

Program / web server	Supplier	Application	Link
Bioedit 7.0.5.3	Tom Hall Ibis Therapeutics	Alignment analysis	
BLASTP	NCBI	Protein database mining	https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome
ChemOffice 2016	PerkinElmer	Chemical structures and calculations	

Program / web server	Supplier	Application	Link
EMBOSS Needle	EMBL-EBI	Pairwise sequence alignment	https://www.ebi.ac.uk/Tools/psa/emboss_needle/
IQ-TREE	IQ-TREE Google group	Phylogenetic tree	http://www.iqtree.org/
LabSolution	Shimadzu	HPLC analysis	
MAFFT V7	CBRC	Multi sequence alignment, phylogenetic tree	https://mafft.cbrc.jp/alignment/server/
MS Office 2016	Microsoft	Text and table calculations	
OriginLab 2018	OriginLab corporation	Graphic and functional analysis	
PPR 1	P.K. Busk, L. Lange, Aalborg University	Peptide pattern recognition	https://vbn.aau.dk/en/publications/peptidm%C3%B8nstergenkendelse
Serial Cloner 2.6	Serial Basics	Sequencing analysis	
TopSpin 4.0.7	Bruker	NMR analysis	
webPRANK	EMBL-EBI	Multi sequence alignment	https://www.ebi.ac.uk/goldman-srv/webprank/

2.2.1 Bacterial strains and plasmids

Table 2.3 Bacterial strains used in this thesis. The sources of the used bacterial strains are Thermo Fisher Scientific (Waltham, USA-MA) and Agilent Technologies (Waldbronn, Germany).

Strain	Purpose	Genotype	Source
E. coli DH5α	Molecular biology	F- ϕ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-, mk+) gal- phoA supE44 λ- thi-1 gyrA96 relA1	Thermo Fisher Scientific
E. coli XL1-Blue	Molecular biology	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proAB lacIq ZΔM15 Tn10 (Tetr)]	Agilent Technologies
E. coli BL21 (DE3) Gold	Protein expression	E. coli B F– ompT hsdS(rB – mB –) dcm+ Tetr gal λ(DE3) endA Hte	Agilent Technologies

2.2.2 Primer

Table 2.4 Primer used in this thesis. The primer were synthesized by Sigma-Aldrich (München, Germany).

Primer	Sequence
T7 FW	5'-taatacgactcactataggg-3'
T7 RV	5'-gctagttattgctcagcgg-3'

Primer	Sequence
Quikchange Primer	
LigG-TD C12S FW	5'-gtttaccacattccggtctccccgttctgc-3'
LigG-TD C12S RV	5'-gcagaacggggagaccggaatgtggtaaac-3'
LigG-TD C15S FW	5'-tcaacacgttgggagaacgggcagaccg-3'
LigG-TD C15S RV	5'-cggtctgcccgttctcccaacgtgttga-3
LigG-TD A55S FW	5'-ccagcaccggcagggaggtcgtaccacgc-3'
LigG-TD A55S RV	5'-gcgtggtacgacctccctgccggtgctgg-3'
LigG-TD D104A FW	5'-gcaacgaaatcacgggccatggtggtcagca-3'
LigG-TD D104A RV	5'-tgctgaccaccatggcccgtgatttcgttgc-3'
LigG-TD D104E FW	5'-gcaacgaaatcacgctccatggtggtcagca-3'
LigG-TD D104E RV	5'-tgctgaccaccatggagcgtgatttcgttgc-3'
LigG-TD R105A FW	5'-gccgcaacgaaatcagcgtccatggtggtcag-3'
LigG-TD R105A RV	5'-ctgaccaccatggacgctgatttcgttgcggc-3'
LigG-TD V108A FW	5'-catggaccgtgatttcgctgcggcgggtta-3'
LigG-TD V108A RV	5'-taacccgccgcagcgaaatcacggtccatg-3'
LigG-TD V108Y Y112A FW	5'-tcatcagccaaccgccacccgccgcatagaaatcacggtccatg-3'
LigG-TD V108Y Y112A RV	5'-catggaccgtgatttctatgcggcgggtggcggttggctgatga-3'
LigG-TD Y112A FW	5'-tcatcagccaaccgccacccgccgcaacgaaatc-3'
LigG-TD Y112A RV	5'-gatttcgttgcggcgggtggcggttggctgatga-3'
LigG-TD Y112F FW	5'-cagccaaccgaaacccgccgcaacg-3'
LigG-TD Y112F RV	5'-cgttgcggcgggtttcggttggctg-3'
LigG-TD Y112W FW	5'-gttcatcagccaaccccaacccgccgcaacgaaat-3'
LigG-TD Y112W RV	5'-atttcgttgcggcgggttggggttggctgatgaac-3'
LigG-TD M116A FW	5'-gtttcggatcttggttcgccagccaaccgtaacccg-3'
LigG-TD M116A RV	5'-cgggttacggttggctggcgaaccaagatccgaaac-3'
LigG-TD F165T FW	5'-aaccagaaacgctggaaggccggggtgaaaacggtttc-3'
LigG-TD F165T RV	5'-gaaccagaaacgctggaacatcggggtgaaaacggtttc-3'
Saturation mutagenesis primer	
LigG-TD V11 NDT FW	5'-ndttgcccgttctgccaacgtgttgaaatcc-3'
LigG-TD V11 VMA FW	5'-vmatgcccgttctgccaacgtgttgaaatcc-3'
LigG-TD V11 ATG FW	5'-atgtgcccgttctgccaacgtgttgaaatcc-3'
LigG-TD V11 TGG FW	5'-tggtgcccgttctgccaacgtgttgaaatcc-3'
LigG-TD V11 RV	5'-cggaatgtggtaaacggtcggacgggtcatatg-3'
LigG-TD A55 NDT FW	5'-accndtctgccggtgctggagaccgcg-3'
LigG-TD A55 VMA FW	5'-accvmactgccggtgctggagaccgcg-3'
LigG-TD A55 ATG FW	5'-accatgctgccggtgctggagaccgcg-3'
LigG-TD A55 TGG FW	5'-acctggctgccggtgctggagaccgcg-3'
LigG-TD A55 RV	5'-cgtaccacgcgtcttttgcagcagccagtccggacg-3'
LigG-TD D104 NDT FW	5'-ndtcgtgatttcgttgcggcgggttacggttggc-3'
LigG-TD D104 VMA FW	5'-vmatcgtgatttcgttgcggcgggttacggttggc-3'
LigG-TD D104 ATG FW	5'-atgcgtgatttcgttgcggcgggttacggttggc-3'
LigG-TD D104 TGG FW	5'-tggtcgtgatttcgttgcggcgggttacggttggc-3'
LigG-TD D104 RV	5'-catggtggtcagcatattctccaccgcacgacggtacggatc-3'
LigG-TD R105 NDT FW	5'-ndtgatttcgttgcggcgggttacggttggctgatg-3'

Primer	Sequence
LigG-TD R105 VMA FW	5'-vmagatttcgttgcggcgggttacggttggctgatg-3'
LigG-TD R105 ATG FW	5'-atggatttcgttgcggcgggttacggttggctgatg-3'
LigG-TD R105 TGG FW	5'-tgggatttcgttgcggcgggttacggttggctgatg-3'
LigG-TD R105 RV	5'-gtccatggtggtcagcatattctccaccgcacgacgg-3'
LigG-TD V108 NDT FW	5'-ndtgcggcgggttacggttggctgatgaacc-3'
LigG-TD V108 VMA FW	5'-vmagcggcgggttacggttggctgatgaacc-3'
LigG-TD V108 ATG FW	5'-atggcggcgggttacggttggctgatgaacc-3'
LigG-TD V108 TGG FW	5'-tgggcggcgggttacggttggctgatgaacc-3'
LigG-TD V108 RV	5'-gaaatcacggtccatggtggtcagcatattctccaccgc-3'
LigG-TD M116 NDT FW	5'-ndtaaccaagatccgaaacaacgtgatgcgctgcgt-3'
LigG-TD M116 VMA FW	5'-vmaaaccaagatccgaaacaacgtgatgcgctgcgt-3'
LigG-TD M116 ATG FW	5'-atgaaccaagatccgaaacaacgtgatgcgctgcgt-3'
LigG-TD M116 TGG FW	5'-tggaaccaagatccgaaacaacgtgatgcgctgcgt-3'
LigG-TD M116 RV	5'-cagccaaccgtaacccgccgcaacgaaatcacg-3'
LigG-TD F165 NDT FW	5'-ndtttccagcgtttctggttcctggagtactacgaggatttcg-3'
LigG-TD F165 VMA FW	5'-vmattccagcgtttctggttcctggagtactacgaggatttcg-3'
LigG-TD F165 ATG FW	5'-atgttccagcgtttctggttcctggagtactacgaggatttcg-3'
LigG-TD F165 TGG FW	5'-tggttccagcgtttctggttcctggagtactacgaggatttcg-3'
LigG-TD F165 RV	5'-cggggtgaaaacggtttcagcccaaccgaaggtttcg-3'
LigG-TD N223 NDT FW	5'-ndtggtgcgctgctgccaggtcgtagc-3'
LigG-TD N223 VMA FW	5'-vmaggtgcgctgctgccaggtcgtagc-3'
LigG-TD N223 ATG FW	5'-atgggtgcgctgctgccaggtcgtagc-3'
LigG-TD N223 TGG FW	5'-tggggtgcgctgctgccaggtcgtagc-3'
LigG-TD N223 RV	5'-acccgcaccgcacgcgtaatcatagtacagc-3'
LigG-TD V108 NNS FW	5'-ccgtaacccgccgcsnngaaatcacggtccatggtg-3'
LigG-TD V108 NNS RV	5'-caccatggaccgtgatttcnnsgcggcgggttacgg-3'
LigG-TD M116 NNS FW	5'-gttgtttcggatcttggttshncagccaaccgtaacccgcc-3'
LigG-TD M116 NNS RV	5'-ggcgggttacggttggctgndsaaccaagatccgaaacaac-3'
LigG-TD N223 NNS FW	5'-gcgtgcggtgcgggtndsggtgcgctgctgc-3'
LigG-TD N223 NNS RV	5'-gcagcagcgcaccshnacccgcaccgcacgc-3'

2.2.3 Antibiotics

The antibiotics kanamycin and ampicillin were used as 1000 times stock solutions. The concentrations of the stock solutions were 50 $mg \cdot mL^{-1}$ for kanamycin and 100 $mg \cdot mL^{-1}$ for ampicillin. The antibiotics were solved in water, sterile filtered and stored at -20 °C.

2.2.4 Buffer and media

Table 2.5 Media and buffers used in this thesis with their composition.

Buffer/Media	Components	Remarks
LB media (Lysogeny broth)	10 $g \cdot L^{-1}$ tryptone 5 $g \cdot L^{-1}$ yeast extract 10 $g \cdot L^{-1}$ NaCl	autoclaved
LB agar	LB media + 15 $g \cdot L^{-1}$ agar	
TB media (Terrific Broth)	12 $g \cdot L^{-1}$ tryptone 24 $g \cdot L^{-1}$ yeast extract 5 $g \cdot L^{-1}$ glycerol	-add 900 mL MPW; -autoclave; -before usage add 100 mL 10x TB salt solution
10x TB salts solution	23.1 $g \cdot L^{-1}$ KH_2PO_4 125.4 $g \cdot L^{-1}$ K_2HPO_4	autoclave
SOC	20 $g \cdot L^{-1}$ tryptone 5 $g \cdot L^{-1}$ yeast extract 0.5 $g \cdot L^{-1}$ NaCl 0.186 $g \cdot L^{-1}$ KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ 20 mM glucose	-prepare solution of tryptone, yeast extract, NaCl, and KCl; autoclave -prepare separate solution of 2 M $MgCl_2$, 2 M $MgSO_4$, 2 M glucose -sterilize $MgCl_2$ and $MgSO_4$ by autoclaving and glucose by filtration -add $MgCl_2$ and $MgSO_4$ (1/200) and glucose (1/100)
IMAC binding buffer A	20 mM KH_2PO_4 500 mM NaCl 20 mM imidazole	adjust pH with HCl to pH 7.4; filter and degas the buffer
IMAC binding buffer B	20 mM KH_2PO_4 500 mM NaCl 500 mM imidazole	adjust pH with HCl to pH 7.4; filter and degas the buffer
Protein storage buffer	20 mM Tris 20% glycerol	adjust pH with HCl to pH 7.5; autoclave
50x TAE	242 $g \cdot L^{-1}$ Tris 57.1 $mL \cdot L^{-1}$ acetic acid 100 $mL \cdot L^{-1}$ 0.5 M EDTA pH 8	
10x SDS	288 $g \cdot L^{-1}$ glycine 20 $g \cdot L^{-1}$ SDS 60 $g \cdot L^{-1}$ Tris	
4x SDS sample buffer	80 $mg \cdot L^{-1}$ SDS 40% glycerol 20% mercaptoethanol 4 $mg \cdot L^{-1}$ bromphenol blue 100 mM Tris/HCl pH 6.8	Prepare solution of SDS in Tris/HCl (stir; no shaking); add other components
SDS staining solution	30% ethanol 10% acetic acid 2.5 $g \cdot L^{-1}$ Coomassie-Brillant-Blue	
SDS destaining solution	30% ethanol 10% acetic acid	

2.2.5 Proteins

Table 2.6 Proteins expressed in this study. As expression vector in all cases pET28a with the N-terminal his_6-tag was used. The genes were cloned using the restriction enzymes NdeI and HindIII. The biochemical parameters were calculated with the Protparam online server.[91] DNA and protein sequence of the expressed proteins are listed in section 6.1.

Enzyme	Origin	GenBank accession number	Amino acids	Molecular weight [Da]	pI	Absorbance factor 1/ε [g·L^{-1}·cm]	Reference
LigE	*Sphingobium* sp. SYK-6	WP_014075192	301	34234	5.81	0.463	[30]
LigE-NS	*Novosphingobium* sp. PP1Y	WP_013832481	295	33210	5.83	0.527	[30]
LigE-NA	*N. aromaticivorans* DSM 12444	WP_011446047	299	33286	6.22	0.622	[30]
LigP	*Sphingobium* sp. SYK-6	WP_014077574	301	33160	6.26	0.593	[30]
LigE179	*Altererythrobacter atlanticus*	WP_046903179	301	33328	5.88	0.542	
LigE283	*Altererythrobacter* sp. 66-12	OJU60283	300	33389	6.17	0.566	
LigE491	*Sphingomonas hengshuiensis*	WP_044331491	296	33141	6.04	0.503	
LigE760	*Novosphingobium* sp. SCN 63-17	ODU84760	303	33842	6.18	0.537	
LigE889	*Altererythrobacter* sp. Root672	WP_055920889	300	33322	5.78	0.565	
LigE915	*Novosphingobium capsulatum*	WP_062781915	301	33105	6.38	0.561	
LigF	*Sphingobium sp.* SYK-6	WP_014075191	277	31849	6.38	0.584	[30]
LigF-NS	*Novosphingobium* sp. PP1Y	WP_013832480	270	30755	6.19	0.614	[30]
LigF-NA	*N. aromaticivorans* DSM 12444	WP_041551020	266	30739	6.06	0.563	[30]
LigF008	*Altererythrobacter* sp. Root672	WP_055919008	266	30668	6.31	0.578	
LigF215	Gammaproteobateria	OGT78215	260	30251	6.84	0.526	
LigF729	*Novosphingobium* sp. SCN 63-17	ODU83729	266	30621	6.09	0.561	
LigF755	*Sphingobium* sp. TCM1	WP_066854755	271	30753	6.06	0.579	
LigF921	*Erythrobacter* sp. SG61-1L	WP_054529921	268	31527	6.01	0.470	
LigF935	*Altererythrobacter* sp. 66-12	OJU59935	271	31224	6.46	0.602	
LigF965	*Novosphingobium lentum*	WP_068075965	266	30319	6.18	0.541	
LigG	*Sphingobium* sp. SYK-6	BAA77216	285	32443	5.96	0.637	[59]
LigG-NS	*Novosphingobium* sp. PP1Y	CCA92089	297	34053	6.09	0.603	[59]
LigG-TD	*T. denitrificans* ATCC 25259	WP_011311562	283	32855	637	0.567	[59]
GST3	*Novosphingobium* sp. MBES04	WP_039391121	246	28158	5.76	0.531	[63,68]
LigG817	*Novosphingobium* sp. PP1Y	WP_041558817	246	28004	5.99	0.602	

Table 2.7 Proteins and enzymes kits from commercial distributers. The sources of the used bacterial strains are New England BioLabs (Frankfurt am Main, Germany), Thermo Fisher Scientific (Waltham, USA-MA), Agilent Technologies (Waldbronn, Germany), and Sigma-Aldrich (München, Germany).

Proteins	Concentration	Distributer
NdeI	20.000 U·mL^{-1}	New England BioLabs
HindIII	20.000 U·mL^{-1}	New England BioLabs
DpnI	20.000 U·mL^{-1}	New England BioLabs
T4-ligase	400.000 U·mL^{-1}	New England BioLabs
T4 polynucleotide kinase	10.000 U·mL^{-1}	New England BioLabs
DreamTaq		Thermo Fischer Scientific
Q5® High-Fidelity 2X Master Mix		New England BioLabs
PfuUltra II Hotstart PCR Master Mix		Agilent Technologies
Glutathione reductase from baker's yeast		Sigma-Aldrich

2.2.6 Kits and standards

Table 2.8 Kits and standards used in this study.

Kit/standard	Purpose	Producer
E.Z.N.A.® Plasmid Mini Kit I	Plasmid extraction	Omega Bio-tek (Norcross, USA-GA)
E.Z.N.A.® MicroElute DNA Clean-Up Kit	PCR/restriction purification	Omega Bio-tek (Norcross, USA-GA)
GeneRuler DNA Ladder Mix	DNA molecular weight marker	Thermo Fisher Scientific (Waltham, USA-MA)
Pierce™ Unstained Protein Molecular Weight Marker	Protein molecular weight marker	Thermo Fisher Scientific (Waltham, USA-MA)

2.3 Methods

2.3.1 Substrate synthesis

The substrate synthesis was performed after modified protocols used by Picart et al.[30] The three step synthesis starts with the bromination of a acetophenone derivate (section 2.3.1.1). The keto-ether bond is formed in a S_N2 reaction with a phenol compound as nucleophile (section 2.3.1.2). The last step is the hydroxy methylation reaction of the keto-ether compound with formaldehyde (section 2.3.1.3). ^{1}H NMR spectra of all synthesized compounds are shown in section 6.2.

The substrate synthesis strategy is summarized in Figure 2.1, while the names of the substrates are listed in Table 2.9

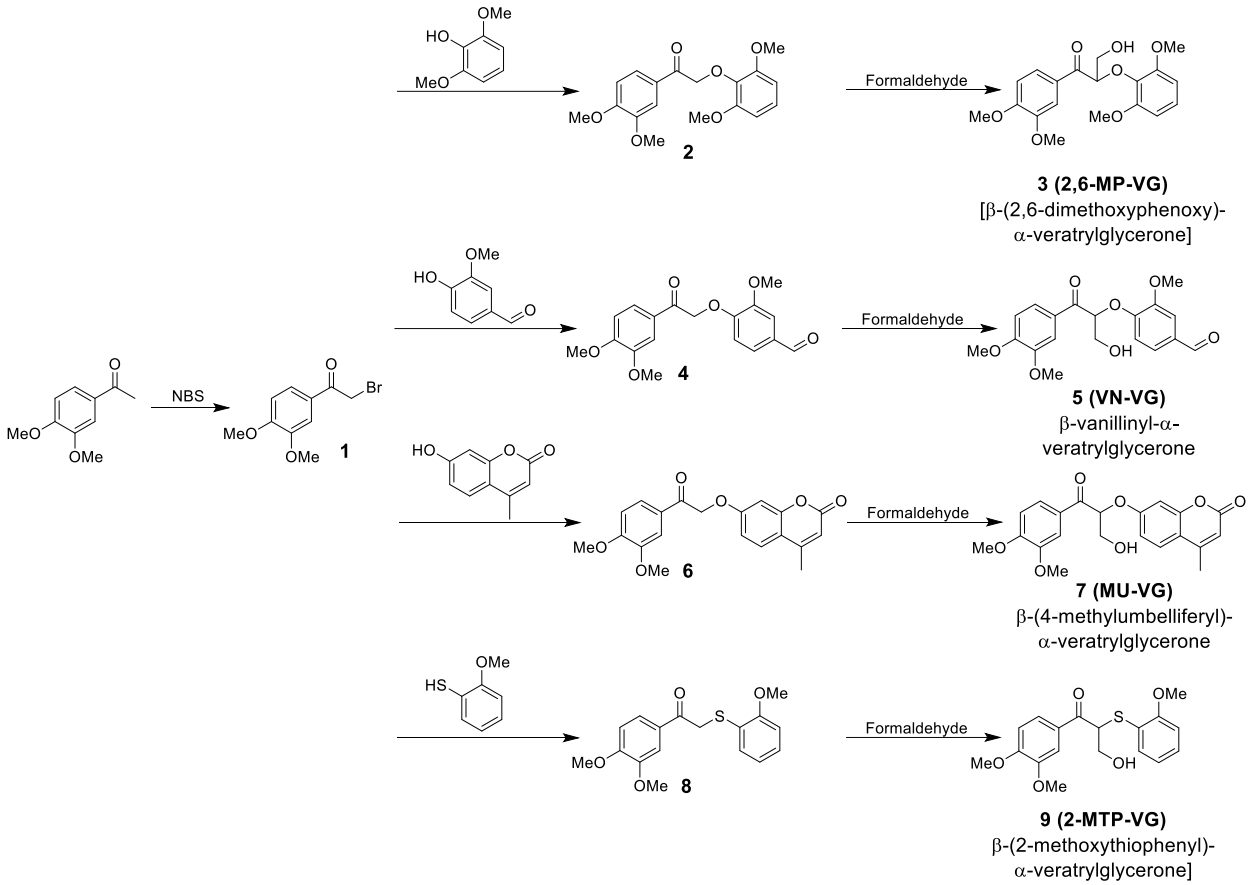


Figure 2.1 General scheme of the substrate synthesis. Reaction steps are bromination, ether formation and hydroxy methylation. The systematic names of the products are listed in Table 2.9.

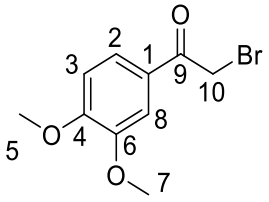
Table 2.9 References and names of synthesized substrates.

Compound	Reference	Full-Name	Short Name
1		2-bromo-1-(3,4-dimethoxyphenyl)ethan-1-one	-
2		2-(2,6-dimethoxyphenoxy)-1-(3,4-dimethoxyphenyl)ethan-1-one	-
3	Picart, 2014[30]	2-(2,6-dimethoxyphenoxy)-1-(3,4-dimethoxyphenyl)-3-hydroxypropan-1-one	2,6-MP-VG
4		4-(2-(3,4-dimethoxyphenyl)-2-oxoethoxy)-3-methoxybenzaldehyde	-
5	Helmich, 2016[69]	4-((1-(3,4-dimethoxyphenyl)-3-hydroxy-1-oxopropan-2-yl)oxy)-3-methoxybenzaldehyde	VN-VG
6		7-(2-(3,4-dimethoxyphenyl)-2-oxoethoxy)-4-methyl-2H-chromen-2-one	-
7		7-((1-(3,4-dimethoxyphenyl)-3-hydroxy-1-oxopropan-2-yl)oxy)-4-methyl-2H-chromen-2-one	MU-VG
8		1-(3,4-dimethoxyphenyl)-2-((2-methoxyphenyl)thio)ethan-1-one	
9		2-(2,6-dimethoxyphenoxy)-1-(3,4-dimethoxyphenyl)-3-hydroxypropan-1-one	

2.3.1.1 Synthesis of 2-bromo-1-(3,4-dimethoxyphenyl)ethan-1-one (1)

The reaction was carried out under Schlenk conditions in a three neck-flask with reflux condenser. The bromintation reagent *N*-bromsuccinimid (NBS) is water sensitive, therefore water needs to be excluded from the reaction.

The solvent acetonitrile was dried overnight on activated molecular sieve (3 Å Pore size). 4.59 g 3,4-dimethoxyacetophenon (25.47 mmol) and 6.6 g *p*-toluene sulfonic acid monohydrate (38.2 mmol, 1.5 eq.) were dissolved in 400 mL acetonitrile. NBS (4.98 g, 28 mmol, 1.1 eq.) was dissolved in 100 mL acetonitrile and then added to the reaction mixture. The reaction mixture was stirred and heated to 110 °C for 2 h. The solvent was evaporated. The crude product was separated between H_2O and DCM, the aqueous phase extracted three times with DCM and the solvent of combined organic phase removed. The product was purified with DCM as mobile phase. To separate product and by-product (α-dibromo-3,4-dimethoxyacetophenon) this step was performed 2 to 3 times. The product **1** was isolated as slightly purple solid with a yield of 80.2% (5.29 g).



^{1}H NMR (400 MHz, CCl_3D): δ [ppm] = 7.61 (dd, ^{3}J = 8.3 Hz, ^{5}J = 2.0 Hz, 2-C**H**, 1H), 7.53 (d, ^{5}J = 2.0 Hz, 8-C**H**, 1H), 6.90 (d, ^{3}J = 8.3 Hz, 3-C**H**, 1H), 4.41 (s, 10-C**H**, 2H), 3.96 (s, 7-OC**H**$_3$, 3H), 3.94 (s, 5-OC**H**$_3$, 3H)

^{13}C NMR (100 MHz, CCl_3D): δ [ppm] = 190.2 (9-**C**O), 154.2 (6-**C**), 149.5 (4-**C**), 127.2 (1-**C**), 124.0 (2-**C**), 111.0 (8-**C**), 110.3 (3-**C**), 56.3 (7-0**C**), 56.2 (5-O**C**), 30.5 (10-**C**)

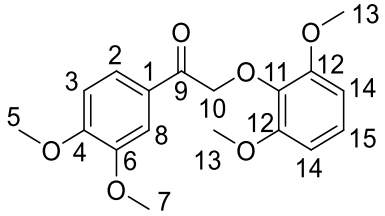
ESI MS: $[M+H^+]^+$ = 258.97 m/z (calculated 259.00)

2.3.1.2 Synthesis of β-keto-ethers (2, 4, 6, 8)

Product **1** was dissolved with 1.5 eq. DIPEA (*N,N*-diisopropylethylamine) and phenol reagent (**2**: 2,6-dimethoxyphenol, **4**: vanillin, **6**: 4-methylumbiliferone, **8**: 2-methoxy-thiophenole) in acetone and stirred at room temperature. After the end of the reaction the solvent was removed, and the crude product purified by column chromatography.

Synthesis of 2

The crude product was purified by column chromatography with gradient of DCM to 20% EtOAc in DCM. The product **2** was isolated with a yield of 79.4%.



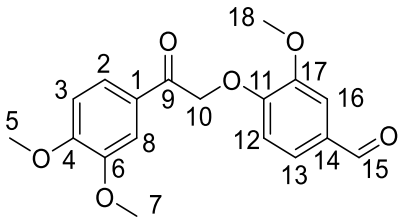
^{1}H NMR (600 MHz, CCl_3D): δ [ppm] =7.71 (dd, ^{3}J = 8.5 Hz ^{5}J = 2.0 Hz, 2-C**H**, 1H), 7.50 (d, ^{5}J = 2.0 Hz, 8-C**H**, 1H), 7.07 (d, ^{3}J = 8.5 Hz, 3-C**H**, 1H), 7.02 (t, ^{3}J = 8.4 Hz, 15-C**H** 1H) 6.68 (d, 3J = 8.4 Hz, 14-C**H**, 2H), 5.08 (s, 10-C**H**$_2$, 2H), 3.85 (s, 5-OC**H**$_3$, 3H), 3.82 (s, 7-OC**H**$_3$, 3H), 3.74 (s, 13-OC**H**$_3$, 6H)

^{13}C NMR (150 MHz, CCl_3D): δ [ppm] = 193.2 (9-**C**O), 153.3 (4-**C**), 152.9 (12-**C**), 148 (6-**C**), 135.9 (11-**C**O), 127.6 (1-**C**), 124.0 (15-**C**), 122.8 (2-**C**), 110.9 (3-**C**), 110.3 (8-**C**), 105.6 (14-**C**), 74.5 (10-**C**), 55.9 (13-0**C**), 55.8 (5-O**C**), 55.5 (7-O**C**)

ESI MS: $[M+H^+]^+$ = 333.00 m/z (calculated 333.13)

Synthesis of 4

The crude product was purified by column chromatography with gradient of pure DCM to 20% EtOAc in DCM. The product **4** was isolated with a yield of 93.4%.



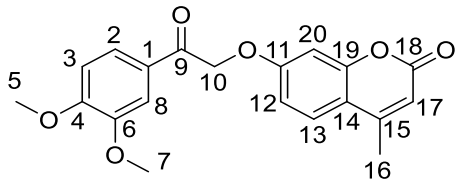
^{1}H NMR (400 MHz, DMSO): δ [ppm] = 9.83 (s, 15-C**H**, 1H) 7.72 (dd, ^{3}J = 8.3 Hz, ^{5}J = 2.0 Hz, 2-C**H**, 1H), 7.49 (d, ^{5}J = 2.0 Hz, 8-C**H**, 1H), 7.48 (dd, ^{3}J = 8.6 Hz, ^{5}J = 1.9 Hz, 13-C**H**, 1H), 7.43 (d, ^{5}J = 1.9 Hz, 16-C**H**, 1H), 7.12 (d, ^{3}J = 8.3 Hz, 3-C**H**, 1H), 7.05 (d, ^{3}J = 8.6 Hz, 12-C**H**, 1H), 5.69 (s, 10-C**H**2, 2H), 3.87 (s, 18-OC**H**3, 3H), 3.86 (s, 7-OC**H**3, 3H), 3.83 (s, 5-OC**H**3, 3H)

^{13}C NMR (100 MHz, DMSO): δ [ppm] = 192.1 (9-**C**O), 191.3 (16-**C**O), 153.6 (6-**C**), 152.9 (11-**C**O), 149.2 (17-**C**), 148.7 (4-**C**), 129.9 (14-**C**), 127.0 (1-**C**), 125.5 (13-**C**), 122.6 (2-**C**), 112.6 (12-**C**), 111.0 (3-**C**), 110.2 (8-**C**)m 110.1 (16-**C**), 70.3 (10-**C**), 55.8 (7-O**C**), 55.6 (18-O**C**), 55.5 (5-O**C**)

ESI MS: $[M+Na^+]^+$ = 353.03 m/z (calculated 353.10)

Synthesis of 6

The crude product was loaded on silica gel and subsequently purified by column chromatography with a gradient of DCM/EtOAc 1/1 to EtOAc. The product **6** was isolated with a yield of 39.3%.



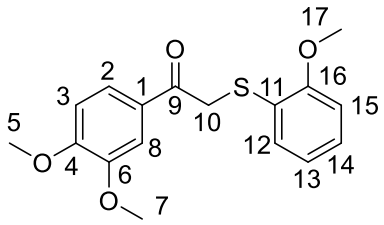
^{1}H NMR (600 MHz, CCl_3D): δ [ppm] = 7.62 (dd, ^{3}J = 8.4 Hz, ^{5}J = 2.0 Hz, 2-C**H**, 1H), 7.54 (d, ^{5}J = 2.0 Hz, 8-C**H**, 1H), 7.51 (d, ^{3}J = 8.8 Hz, 13-C**H**, 1H), 6.95 (dd, ^{3}J = 8.8 Hz, ^{5}J = 2.6 Hz, 12-C**H**, 1H), 6.93 (d ^{3}J = 8.4 Hz, 3-C**H**, 1H), 6.79 (d, ^{5}J = 2.6 Hz, 20-C**H**, 1H), 6.14 (d, ^{4}J = 1.2 Hz, 17-C**H**, 1H), 5.34 (s, 10-C**H**, 2H), 3.98 (s, 5-OC**H**$_3$, 3H), 3.95 (s, 7-OC**H**$_3$, 3H), 2.39 (d, ^{4}J = 1.2 Hz, 16-C**H**$_3$, 3H)

^{13}C NMR (150 MHz, CCl_3D): δ [ppm] = 191.9 (9-**C**O), 161.3 (11-**C**O), 161.1 (18-**C**OO), 155.2 (19-**C**), 154.4 (4-**C**), 152.6 (14-**C**), 149.6 (6-**C**), 127.4 (1-**C**), 125.9 (13-**C**), 122.8 (2-**C**), 114.4 (15-**C**), 112.8 (12-**C**), 112.5 (17-**C**), 110.4 (3-**C**), 110.3 (8-**C**), 102.0 (20-**C**), 70.5 (10-**C**), 56.3 (5-O**C**), 56.2 (7-O**C**), 18.8 (16-**C**)

ESI MS: $[M+H^+]^+$ = 355.18 m/z (calculated 355.12)

Synthesis of 8

The crude product was purified by column chromatography with gradient of DCM/EtOAc 5:1 to DCM/EtOAc 2:1. The product **8** was isolated with a yield of 96.8%.



^{1}H NMR (300 MHz, CCl_3D): δ [ppm] = 7.58 (dd, ^{3}J = 8.5 Hz, ^{5}J = 2.1 Hz, Ar-**H**, 1H), 7.36 (m, Ar-**H**, 1H), 7.25 (m, Ar-**H**, 1H), 6.88 (m, Ar-**H**, 3H), 4.19 (s, 10-C**H**$_2$, 2H), 3.94 (s, OC**H**$_3$, 3H), 3.90 (s, OC**H**$_3$, 3H), 3.86 (s, OC**H**$_3$, 3H)

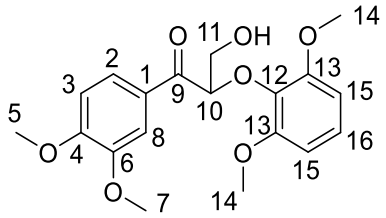
ESI MS: $[M+Na^+]^+$ = 341.21 m/z (calculated 341.08)

2.3.1.3 Hydroxy methylation of β-keto-ethers (3, 5, 7, 9)

The purified product of the β-keto-ether synthesis together with Na_2CO_3 (anhydrous, 0.5 eq.) and paraformaldehyde (10 eq.) were dissolved and stirred at room temperature or 45 °C. The solvent differs depending on the substrate and needs to be optimized for each reaction. The solvent was removed, and the crude product purified by column chromatography.

Synthesis of 3 (2,6-MP-VG)

For the synthesis of 2,6-MP-VG methanol was used as solvent. The crude product was purified by column chromatography with a gradient of DCM/EtOAc from 20/1 to 10/1. The product **3** was isolated with a yield of 84.2%



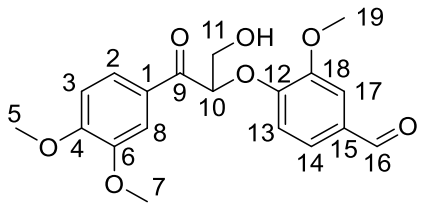
^{1}H NMR (600 MHz, DMSO): δ [ppm] = 7.72 (dd, ^3J = 8.5 Hz, ^5J = 2.1 Hz, 2-C**H**, 1H), 7.50 (d, ^3J = 2.1 Hz, 8-C**H**, 1H), 7.07 (d, ^3J = 8.5 Hz, 3-C**H**, 1H), 6.98 (t, ^3J = 8.3 Hz, 16-C**H**, 1H), 6.54 (d, ^3J = 8.3 Hz, 15-C**H**, 2H), 5.20 (m, 10-C**H**, 1H), 4.73 (m, 11-CH_2O**H**, 1H), 3.84 (s, 5-OC**H**$_3$, 3H), 3.79 (s, 7-OC**H**$_3$, 3H), 3.78 (m, 10-C**H**$_2$, 2H), 3.64 (s, 14-OC**H**$_3$, 3H),

^{13}C NMR (150 MHz, DMSO): δ [ppm] = 195.4 (9-**C**O), 153.1 (4-**C**), 152.5 (13-**C**), 148.5 (6-**C**), 136.0 (12-**C**O), 128.9 (1-**C**), 123.7 (16-**C**), 123.5 (2-**C**), 110.8 (3-**C**, 8-**C**), 105.6 (15-**C**), 83.5 (10-**C**), 62.2 (11-**C**), 55.86 (14-O**C**), 55.8 (5-O**C**), 55.5 (7-O**C**)

ESI MS: $[M+Na^+]^+$ = 385.15 m/z (calculated 385.13)

Synthesis of 5 (VN-VG)

VN-VG was synthesized with isopropanol as solvent. Due to the very slow reaction progress over 1.5 days at room temperature, the reaction was heated to 45 °C for additional 12 h. The solvent was removed, and the crude product was purified in a two-step column purification process. First, the crude product was loaded onto a silica column with a gradient of pure DCM to DCM/EtOAc 1/1 followed by a second column chromatography with C18-silica with a gradient of 1/10 ACN/water to ACN/water 1/1. VN-VG was isolated as slightly yellow solid with a yield of 75.0%.



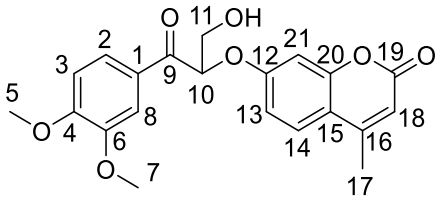
^{1}H NMR (600 MHz, DMSO): δ [ppm] = 9.79 (s, 16-C**H**, 1H), 7.83 (dd, ^{3}J = 8.5 Hz, ^{5}J = 2.0 Hz, 2-C**H**, 1H), 7.53 (d, ^{5}J = 2.0 Hz, 8-C**H**, 1H), 7.42 (m, ^{5}J = 1.9 Hz, 14-C**H**, 1H), 7.41 (m, 17-C**H**, 1 H), 7.11 (d, ^{3}J = 8.5 Hz, 3-C**H**, 1H) 6.89 (d, 8.8 Hz, 13-C**H**, 1H), 5.96 (m, 10-C**H**, 1H), 5.29 (t, ^{3}J = 5.7 Hz, 11-CH_2O**H**, 1H), 3.94 (m, 11-C$\mathbf{H}_2$, 2H), 3.86 (s, 5-OC$\mathbf{H}_3$, 3H), 3.85 (s, 7-OC$\mathbf{H}_3$, 3H), 3.81 (s, 19-OC$\mathbf{H}_3$, 3H)

^{13}C NMR (150 MHz, DMSO): δ [ppm] = 194.0 (9-**C**O), 191.3 (16-**C**O), 153.7 (4-**C**), 152.3 (12-**C**O), 149.3 (18-**C**), 148.7 (6-**C**), 130.0 (15-**C**), 127.6 (1-**C**), 125.5 (14-**C**), 123.4 (2-**C**), 113.1 (13-**C**), 111.0 (3-**C**), 110.7 (8-**C**), 110.4 (17-**C**), 81.4 (10-**C**), 62.5 (11-**C**), 55.9 (5-**C**), 55.6 (19-**C**), 55.5 (7-**C**)

ESI MS: $[M+Na^{+}]^{+}$ = 383.08 m/z (calculated 383.11)

Synthesis of 7 (MU-VG)

For the synthesis of MU-VG THF was used as solvent. The crude product was purified by column chromatography with a gradient of DCM/EtOAc from 1/1 to 10/4. The product **7** was isolated with a yield of 9.2%.



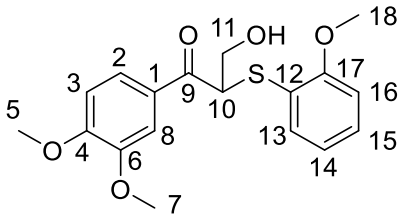
^{1}H NMR (600 MHz, DMSO): δ [ppm] = 7.87 (dd, ^{3}J = 8.5 Hz, ^{5}J = 2 Hz, 2-C**H**, 1H), 7.66 (d, ^{3}J = 8.8 Hz, 14-C**H**, 1H), 7.50 (d, ^{5}J = 2.0 Hz, 8-C**H**, 1H), 7.12 (d, ^{3}J = 8.5 Hz, 3-C**H**, 1H), 6.93 (dd, ^{3}J = 8.8 Hz, ^{5}J = 2.5 Hz, 3-C**H**, 1H), 6.83 (d, ^{5}J = 2.5 Hz, 21-C**H**, 1H), 6.19 (d, ^{4}J = 1.2 Hz, 18-C**H**, 1H), 6.07 (m, 10-C**H**, 1H), 5.32 (t, ^{3}J = 6.0 Hz, 11-CH_2O**H**, 1H), 3.94 (m, 11-C$\mathbf{H}_2$, 2H), 3.87 (s, 5-OC$\mathbf{H}_3$, 3H), 3.82 (s, 7-OC$\mathbf{H}_3$, 3H), 2.37 (s, 17-CH_3, 3H)

^{13}C NMR (150 MHz, DMSO): δ [ppm] = 193.9 (9-**C**O), 160.6 (12-**C**O), 160.0 (19-**C**OO), 154.4 (20-**C**), 153.7 (4-**C**), 153.3 (15-**C**), 148.7 (6-**C**), 127.5 (1-**C**), 126.6 (14-**C**), 123.5 (2-**C**), 113.4 (16-**C**), 112.6 (13-**C**), 111.3 (18-**C**), 111.0 (3-**C**), 110.6 (8-**C**), 101.8 (21-**C**), 80.8 (10-**C**), 62.5 (11-**C**), 55.9 (5-**C**), 55.6 (7-**C**), 18.1 (17-**C**)

ESI MS: $[M+H^{+}]^{+}$ = 385.13 m/z (calculated 385.13)

Synthesis of 9 (2-MTP-VG)

For the synthesis of 2-MTP-VG DMSO was used as solvent. The crude product was purified by column chromatography with DCM as solvent. The product **9** was isolated with a yield of 41.9%.



^{1}H NMR (600 MHz, CCl_3D): δ [ppm] = 7.69 (dd, ^{3}J = 8.4 Hz, ^{5}J = 2.1 Hz, 2-C**H**, 1H), 7.41 (d, ^{3}J = 2.1 Hz, 8-C**H**, 1H), 7.40 (dd, ^{3}J = 7.7 Hz, ^{5}J = 1.8 Hz, 15-C**H**, 1H), 7.31 (m, 13-C**H**, 1H), 7.08 (d, ^{3}J = 8.4 Hz, 3-C**H**, 1H), 7.03 (m, 16-C**H**, 1H), 6.92 (m, 14-C**H**, 1H), 4.88 (q, 3J = 4.6 Hz, 10-C**H**, 1H), 3.92-3.69 (m, 11-C**H**$_2$, 2H), 3.85 (s, 7-OC**H**$_3$, 3H), 3.76 (s, 5-OC**H**$_3$, 3H), 3.75 (s, 18-OC**H**$_3$, 3H)

^{13}C NMR (150 MHz, CCl_3D): δ [ppm] = 194.7 (9-**C**O), 158.4 (17-**C**), 153.3 (6-**C**), 148.6 (4-**C**), 133.3 (15-**C**), 129.6 (13-**C**), 129.1 (1-**C**), 123.3 (2-**C**), 120.9 (12-**C**S), 120.8 (14-**C**), 111.5 (16-**C**), 110.9 (3-**C**), 110.5 (8-**C**), 61.9 (11-**C**), 55.8 (7-O**C**), 55.7 (18-O**C**), 55.3 (5-O**C**), 49.7 (10-**C**)

ESI MS: $[M+Na^{+}]^{+}$ = 371.08 m/z (calculated 371.09)

2.3.1.4 Inhibitor synthesis

The inhibitor synthesis was performed as a three-step synthesis (Figure 2.2) via an aldol condensation with subsequent reduction of the double bond, and final hydroxy methylation reaction. No successful reaction condition could be found for the hydroxy methylation reaction. The aldol condensation and the reduction were performed after protocols of Kausar et al. and Mirza-Aghayan et al.[92,93]

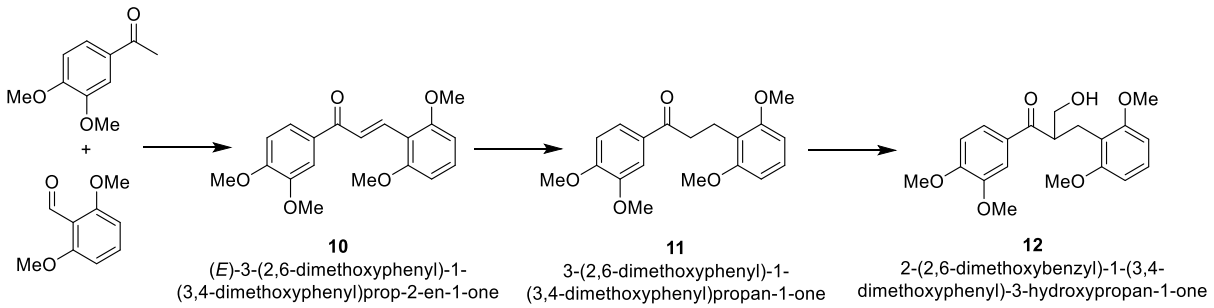
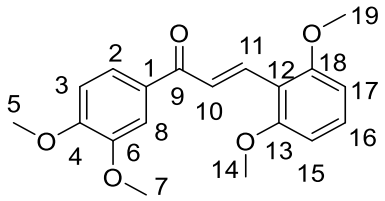


Figure 2.2 Desired synthesis route for the β-etherase inhibitor. Reaction steps are the aldol condensation, reduction of double bond and the hydroxy methylation. Also shown is the systematic name of the products.

Aldol condensation

3,4-dimethoxy-acetophenon (814.8 mg, 4.522 mmol) and 2,6-dimethoxy-benzylaldehyde (751.4 mg, 4.522 mmol) were solved in ethanol and catalytic amounts of NaOH were added. The mixture was stirred overnight at room temperature. The formed product **10** is poorly soluble in ethanol and precipitates during the reaction. The solvent was concentrated, and **10** recrystallized out of ethanol. The solid was washed with ethanol and pentane. **10** was obtained as white needles in a yield of 95.6% (1.420 g, 4.325 mmol).

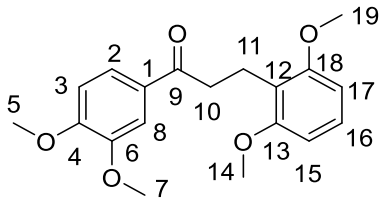


^{1}H NMR (300 MHz, CCl_3D): δ [ppm] = 8.25 (d, ^{3}J = 15.9 Hz, 10-C**H**, 1H), 8.00 (d, ^{3}J = 15.9 Hz, 11-C**H**, 1H), 7.67 (m, 2-C**H**, 8-C**H**, 2 H), 7.29 (t, ^{3}J = 8.4 Hz, 16-C**H**, 1H), 6.93 (d, ^{3}J = 8.4 Hz, 3-C**H** 1H), 6.60 (d, ^{3}J = 8.4 Hz, 15-C**H**, 17-C**H**, 2H), 3.97 (s, 7-OC**H**$_3$, 3H), 3.96 (s, 5-OC**H**$_3$, 3H), 3.92 (s, 14-OC**H**$_3$, 18-OC**H**$_3$, 6H)

ESI MS: $[M+H^+]^+$ = 329.05 m/z (calculated 329.14)

Reduction

10 (0.5469 g, 1.667 mmol) was solved in methanol and EtOAc (1/1). The solution was degassed and saturated with argon atmosphere. 10 wt% Pd/C were added under argon stream. Triethyl silane (10 eq.) was dropped into the reaction mixture over 30 min. After additional 30 min of reaction, the solvent was evaporated, and the crude product was purified over silica with DCM as solvent. **11** was obtained with a yield of 97.6% (537.5 mg, 1.627 mmol).



^{1}H NMR (300 MHz, CCl_3D): δ [ppm] = 7.12 (t, ^{3}J = 8.3 Hz, 16-C**H**, , 1 H), 6.76 (m, 2-C**H**, 3-C**H**, 8-C**H**, 3H), 6.54 (d, ^{3}J = 8.4 Hz, 15-C**H**, 17-C**H**, 2H), 3.86 (s, 7-OC**H**$_3$, 3H), 3.86 (s, 5-OC**H**$_3$, 3H), 3.80 (s, 14-OC**H**$_3$, 18-OC**H**$_3$, 6H), 2.70 (t, 3J = 7.6 Hz, 10-C**H**$_2$, 2H), 2.62 (t, 3J = 7.6 Hz, 11-C**H**$_2$, 2H)

2.3.2 GSH specific dye L_2 synthesis

The synthesis of L2 was performed after a protocol of Das et al (Figure 2.3).[94]

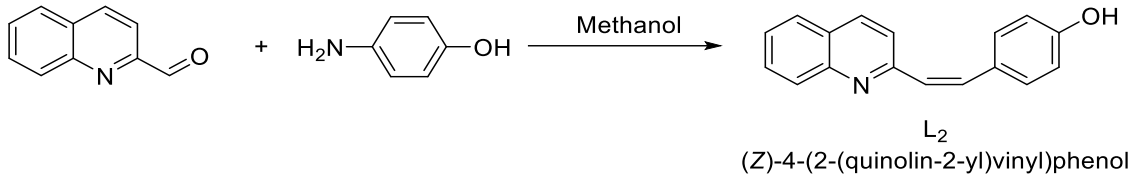
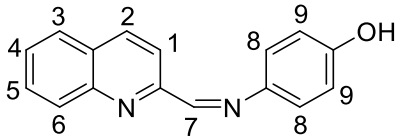


Figure 2.3 Reaction scheme of the GSH specific dye L_2.

2-quinolinecarboxaldehyde (250 mg, 1.59 mmol) was dissolved in 20 mL methanol. *p*-aminophenol (174 mg, 1.59 mmol) was added to the solution. The mixture was stirred overnight, and the product precipitated during the reaction. The product was filtered and washed with cooled methanol. L_2 was obtained in a yield of 78.6% (310 mg, 78.6%).



^{1}H NMR (300 MHz, DMSO): δ [ppm] = 8.80 (s, 7-C**H**, 1 H), 8.48 (d, ^3J = 8.8 Hz, 2-C**H**, 1H), 8.29 (d, ^3J = 8.8 Hz, 1-C**H**, 1H), 8.09 (q, ^3J = 8.4 Hz, 4-C**H**, 5-C**H**, 2H) 7.84 (m, ^3J = 8.4 Hz, 3-C**H**, 1H), 7.84 (m, ^3J = 8.4 Hz, 6-C**H**, 1H), 7.69 (m, ^3J = 8.4 Hz, 6-C**H**, 1H), 7.41 (d, ^3J = 8.9 Hz, 8-C**H**, 1H), 6.87 (d, ^3J = 8.9 Hz, 9-C**H**, 1H)

ESI MS: $[M+Na^+]^+$ = 249.23 m/z (calculated 249.10)

2.3.3 Bioinformatical Methods

2.3.3.1 BLASTP and PHI-BLAST

BLASTP (Basic Local Alignment Search Tool, Protein BLAST) is an alignment based online tool provided by NCBI to search big sequence database. A protein sequence is used as search query and the algorithm creates short sequence parts (3 AA) of this sequence. The database is screened by this short sequence parts and aligns them, while the results are scored. By this scoring spots with high sequence identity are identified, so called high-scoring segment pair (HSP). Next the algorithm tries to extend the local alignment to bigger protein sequence parts and scores this by length and identity of the alignment.[72]

BLASTP was performed using standard parameters in the nr database of GenBank, with an output size of 500, 1.000 or 5.000 sequences.

2.3.3.2 Peptide pattern recognition

Peptide pattern recognition (PPR) is a bioinformatic tool produced by Busk and Lange to group proteins with the same function.[73] PPR analyses if peptides are conserved within a protein database and groups the proteins due to this conservation.

PPR needs a protein sequence database. This database was created by using BLASTP either using the result of a single run or the combination of serval BLASTP runs with deleting the double entries. The PPR tool was used with the default parameters (peptide length 6, number of peptides 70, cut off 10).

2.3.3.3 Alignments and phylogenetic trees

Sequence alignments are used to compare two or more sequences with each other. Equal or similar parts of the sequences are arranged among themselves. Since the order in the sequences cannot be changed, gaps are included to arrange the sequence with maximum consistency. With multiple sequence alignments conserved amino acids or sequence parts within an enzyme family can be identified. A multiple sequence alignment can be used for the creation of a phylogenetic tree. Based on the alignment the relationship and the evolutionary distance is calculated.

Alignments and phylogenetic trees were created to analyse different datasets of protein sequences further. Small datasets (<150 sequences) were aligned with the web server webPRANK and the distance trees were created with the IQTREE web server using the maximum likelihood method.[95,96] The alignment and the creation of distance trees for bigger datasets were performed with the MAFFT web server (version 7, alignment: FFT-NS-2 algorithm, distance tree: Average linkage algorithm UPGMA).[97]

Logos of conserved amino acid motifs were generated using WebLogo.[98]

2.3.3.4 Docking

Docking is a computational approach to predict if and how a ligand binds to a receptor. Therefore, the structure of the receptor (enzyme) and of the ligand (substrate) are used. The docking algorithm tries to place the ligand in the proposed ligand binding site of the receptor and scores the results by calculating the binding energy.

The docking of small molecules was performed using AutoDock Vina implemented in Yasara with 200 to 999 docking runs.[99] The 3D models of the small molecules were created and energy optimized using Chem3D (ChemDraw 3D).

2.3.4 Molecular biological methods

2.3.4.1 Gene synthesis

All genes were synthesized with codon optimized sequences for expression in *E. coli* by Eurofins Genomics (Ebersberg, Germany) with 5'-NdeI and a 3'-HindIII enzyme restriction sides. The genes were subcloned in pET28a using restrictions enzymes NdeI and HindIII and ligated with T4-ligase.

2.3.4.2 Restriction digest

Restriction endonucleases Typ II are enzymes, which specifically cut DNA sequences within or near to a recognition sequence. Depending on the used enzymes, the resulting DNA ends are sticky or blunt. Sticky ends mean that the DNA has 5'or 3' overhangs, respectively, that can be used for specific ligations. If two different restriction enzymes are used to digest the DNA insert and the same two enzymes for the target vector, the orientation of the insert is fixed and double inserts are excluded.

In Table 2.10 the reaction components for the restriction digest are listed. The digest was performed for 3 h at 37 °C. After 1.5 h fresh restriction enzyme was added.

Table 2.10 Components and their concentrations for restriction digest.

	Volume	End concentration	Addition after 1.5h
CutSmart Buffer (10x)	5 µl	1x	
DNA	variable	1000 ng	
NdeI	1 µL		0.5 µL
HindIII	1 µL		0.5 µL
Filled up with MilliQ to 50 µL			

The digested DNA was purified with the E.Z.N.A.® MicroElute DNA Clean-Up Kit from Omega Bio-tek (Norcross, USA-GA) with an elution volume of 15 µL. The cutting product was analyzed afterwards by agarose gel electrophoresis (section 2.3.8.3).

2.3.4.3 Ligation

In a ligation reaction the ATP-dependent T4-DNA-ligase is used to connect 5'-phosphorylated DNA strains with a 3'-hydroxylgroup of another DNA strand to one new combined DNA strand. The ligation reaction was incubated on ice in the fridge overnight. The reaction components are given in Table 2.11. The insert was used in 5x molar excess compared to the target vector. Since the donor vector backbone was not separated from the insert, the full length of the donor vector (backbone + insert) was used for the calculation. Selection between religation (donor backbone + insert) and ligation (target backbone + insert) was done based on the antibiotic resistance of the target vector (kanamycin resistance). The donor vector contains an ampicillin resistance side and bacteria transformed with the religation product are not able to grow on kanamycin containing media. The reaction mixture of the ligation was directly used for the heat shock transformation (section 2.3.4.4).

Table 2.11 Components and their concentrations for the ligation.

	Volume	**End concentration**
T4 DNA Ligase Buffer (10x)	2 µL	1x
T4 Ligase	1 µL	
Vector pET28	variable	50 ng
Insert	variable	5x molar excess
	Filled up with MilliQ to 20 µL	

2.3.4.4 Heat shock transformation and preparation of chemo competent cells

E. coli is not able to uptake free DNA under normal conditions. In contrast to that, some bacteria are naturally able to uptake free DNA. This is called natural competence. But also *E. coli* is in some preparations competent. One of these preparations are chemo competent cells, which can be used for heat shock transformation. Chemo competent cells are washed and incubated with Ca^{2+} and Rb^{+} ions. The positive charged ions work as a bridge between the negative charged DNA backbone and the negative charged lipopolysaccharides on the bacterial cell membrane and mediate the DNA absorbance to the bacterial cells. The DNA uptake happens through a short heat shock, which makes the membrane permeable.

For the preparation of chemo competent cells 50 mL of LB Media was inoculated with an overnight culture to a starting OD_{600} of 0.1. The culture was incubated at 37 °C with 200 rpm shaking until OD_{600} of 0.4-0.6 was reached. The cells were pelleted (4,000×*g*, 15 min, 4 °C),

resuspended in 20 mL RF1 solution (30 mM KCH_3CO_2, 50 mM $MnCl_2$, 100 mM RbCl, 10 mM $CaCl_2$, 15% glycerol, pH set to 5.8 with acetic acid, filtered sterile) and incubated 30 min on ice. The cells are pelleted again and resuspended in 4 mL RF2 (10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol pH set to 6.8 with NaOH, filtered sterile). The cells are aliquoted in 100 µL portions and directly frozen in liquid nitrogen and stored at -80 °C.

The competent cells were thaw on ice and 1 µL closed plasmid or up to 7 µL ligation reaction (cooled) were added. The cells were mixed smoothly (snapping) and afterwards incubated on ice for 20 min. Following the heat shock for 90 sec at 42 °C, the cells were cooled for 5 min on ice and 500 µL SOC media (prewarmed at 37 °C) was added to the cells and the suspension was incubated at 37 °C for 1 h. The cells were plated on LB-agar plates with appropriate antibiotic and grown afterwards at 37 °C overnight.

2.3.4.5 PCR reactions

The polymerase chain reaction (PCR) is used for the specific in vitro amplification of DNA fragments, which can be used either for analysis or for the preparation of DNA for cloning or mutagenesis. The amplified DNA fragment is defined by small DNA molecules (primer), which bind to the DNA template on both DNA strains and define the start of the DNA-amplification. The amplification process is performed by heat stable DNA-polymerase and consists of three steps. First, the DNA double strains are separated by heating (denaturation). Second, the temperature is reduced, and the primer bind specific to the DNA templet (annealing). Third, the DNA-fragment, framed by the primers, is amplified by the DNA-polymerases (elongation).

Colony PCR

The colony PCR is used to identify clones after a cloning from insert into new vector, which carry the new DNA construct.

Bacteria colonies, which were chosen for testing, were resuspended in 20 µL MilliQ water in one Eppendorf tube and 10 µL is transferred into a second Eppendorf tube and the pipette tip dipped on agar plate. The agar plates were incubated overnight at 37 °C (section 2.3.5.1). One of the two Eppendorf tubes with cell suspension were stored at 4 °C and can be used as inoculum for a liquid culture. The other Eppendorf tube is heated for 10 min at 95 °C and afterwards centrifuged. 1 µL of the supernatant were used as templet for the PCR reaction. The colony PCR reaction was analyzed afterwards by agarose gel electrophoresis (section

2.3.8.3). In Table 2.12 are the reaction components of the colony PCR listed, while Table 2.13 displays the PCR program.

Table 2.12 Components of colony PCR reaction master mix.

	Volume	**End concentration**
Green Go Taq® Flexi Buffer (10x)	2 µL	1x
dNTP Mix (10 mM)	0.4 µL	0.2 mM
T7 Promotor Primer (10 µM)	0.5 µL	0.25 µM
T7 Terminator Primer (10 µM)	0.5 µL	0.25 µM
DreamTag (5 U·µL^{-1})	0.2 µL	1 U
MilliQ	15.4 µL	
19 µL Master mix + 1 µL templet		

Table 2.13 PCR program of colony PCR.

	Temperature	**Time**
Preheating	95 °C	5 sec
Break		
1. Initial denaturation	95 °C	120 sec
2. Denaturation	95 °C	30 sec
3. Annealing	55 °C	30 sec
4. Elongation	72 °C	60 sec·kB^{-1}
5. Final elongation	72 °C	10 min
Step 2-4: 30 cycles		

QuikChange PCR

QuikChange PCR strategy was used for side directed mutagenesis to exchange one target amino acid for another desired one. The mutagenesis is introduced in this method by special designed primers. The primers are complementary to each other and carry the information of the mutation in the middle of the primer. During the PCR reaction the whole vector is amplified and after digestion of the templet the double nicked new construct can be directly transformed in *E. coli*.

For SSM libraries NNS denatured primers were used, resulting in 32 possible codons. For 95% coverage 94 clones must be screened.[100] The primers were designed with the help of the QuikChange Primer Design webtool from Agilent Technologies (https://www.genomics.agilent.com/primerDesignProgram.jsp). In Table 2.14 are the reaction components of the QuikChange PCR listed, while Table 2.15 displays the PCR program.

Table 2.14 Components of QuikChange PCR reaction master mix.

	Volume	End concentration
Pfu Ultra 2 Hoststart 2x Mastermix	25 µL	1x
FW Primer (10 µM)	1 µL	0.5 µM
RV Primer (10 µM)	1 µL	0.5 µM
Template	1 µL	10 ng
MilliQ	22 µL	

Table 2.15 PCR program of Quickchange PCR.

	Temperature	Time
Preheating	95 °C	5 sec
Break		
1. Initial denaturation	95 °C	120 sec
2. Denaturation	95 °C	20 sec
3. Annealing	55 °C	20 sec
4. Elongation	72 °C	15 sec·kB^{-1}
5. Final elongation	72 °C	180 sec
	Step 2-4: 30 cycles	

After the PCR reaction 1 µL DpnI was added and the reaction mixture was incubated for 3 h at 37 °C. After 1.5 h, additional 0.5 µL of DpnI were added. The PCR product was analyzed by agarose gel electrophoresis (section 2.3.8.3) and purified by E.Z.N.A.® MicroElute DNA Clean-Up Kit from Omega Bio-tek (Norcross, USA-GA) with an elution volume of 15 µL.

Q5-mutagenesis

Q5-mutagenesis was used for the creation of SSM libraries as written in the NEB (Frankfurt am Main, Germany) protocol.[88] In contrast to the Quickchange protocol, the primers for the Q5-mutagenesis are not complementary and have their binding regions next to each other. The mutation is only located on one of the primers and can be in middle or 5' end of the primer. Since the binding sides of the primers are next to each other the whole vector is amplified as linear PCR product with blunt ends. The linear DNA is cyclized in KDL (kinase, DpnI, ligase) reaction and the new construct directly transformed in *E. coli*. The cyclization results in much higher transformation efficiency.

For SSM libraries primer mixtures using the Tang trick (NDT, VMA, ATG, TGG)[100] were used, to reduce the codon set to 20 possible codons as result of the mutagenesis. To screen 60

clones is sufficient for 95% coverage.[100] In Table 2.16 are the reaction components of the colony PCR listed, while Table 2.17 displays the PCR program Q5-mutagenesis PCR.

Table 2.16 Components of Q5-mutagenesis PCR reaction.

	Volume	**End concentration**
Q5® Hot Start High-Fidelity 2X Master Mix	12.5 µL	1x
FW Primer (10 µM)	1.25 µL	0.5 µM
RV Primer (10 µM)	1.25 µL	0.5 µM
Template	1 µL	10 ng
MilliQ	9 µL	

Table 2.17PCR program of Q5-mutagenesis PCR.

	Temperature	**Time**
Preheating	98 °C	5 sec
Break		
1. Initial denaturation	95 °C	30 sec
2. Denaturation	95 °C	10 sec
3. Annealing	55 °C	30 sec
4. Elongation	72 °C	30 sec·kB^{-1}
5. Final elongation	72 °C	120 sec
Step 2-4: 30 cycles		

After the PCR reaction the KDL reaction (Table 2.18) was performed for 1 h at room temperature.

Table 2.18 Components of KDL reaction.

	Volume	**End concentration**
T4-Ligase Buffer (10x)	1 µL	1X
T4-Polynucleotide-Kinase 10 U·µL^{-1}	0.5 µL	5 U
DpnI 20 U·µL^{-1}	0.5 µL	10 U
Ligase 400 U·µL^{-1}	0.5 µL	200 U
ddH_2O	5.5 µL	
Q5-PCR product	2 µL	

After completion, the KDL reaction mixture was placed on ice and 100 µL chemo competent *E. coli* XL1-Blue cells were added. The heat shock transformation was performed as described in section 2.3.4.4.

2.3.4.6 Plasmid isolation

The plasmid DNA is amplified by *E. coli* strains. The isolation of the plasmid DNA out of these bacteria cells is based on an alkalic pH shift, which results in the lyses of the cell and the denaturation of DNA molecules. After neutralization the genomic DNA precipitates, while the plasmid DNA remains soluble. The cell debris and the genomic DNA are removed by centrifugation and the plasmid DNA is further purified over small spin columns.

For plasmid isolation the E.Z.N.A.® Plasmid Mini Kit I form Omega Bio-Tek (Norcross, USA-GA) was used following the instructions of the producer. Plasmids were isolated out of liquid *E. coli* cultures in shaking flasks using 6 mL of an overnight culture (section 2.3.5.1).

For mutagenesis libraries the mutated plasmids were transformed into *E. coli* XL1-Blue cells (see section 2.3.7.1) and the cells plated on agar plates. The transformed cells were then washed from the agar plate with 2 times 2 mL LB-media and subsequently used for plasmid isolation.

2.3.4.7 Sequencing

For the analysis of cloning or mutagenesis experiments the insert carrying vector was sequenced using sanger sequencing by Eurofins Genomics (Ebersberg, Germany).

2.3.5 Cultivation of bacteria and enzyme expression

2.3.5.1 Overnight cultures

For overnight cultures, LB or TB media with the appropriate antibiotic was inoculated from cryo culture or agar plate. The culture was incubated overnight at 37 °C with 200 rpm or for 2.5 days at 20 °C with 200 rpm.

2.3.5.2 Cryo cultures

For cryo cultures 1 volume of an overnight culture was mixed with the same volume of 50% glycerol (autoclaved, 25% end concentration). The cryo culture was stored at -80 °C.

2.3.5.3 Expression of β-etherases and glutathione lyases

For the heterological expression in *E. coli* the T7-expression system was used. This system consists of the T7-RNA-polymerase encoded in the genome of *E. coli* BL21 (DE3) strains, and a

pET vector with a specific T7 promotor site in front of the target gene. The expression of the T7-RNA-polymerase and of the target gene is controlled by the lac-repressor. After induction with isopropyl-β-D-thiogalactopyranoside (IPTG) the T7-RNA-polymerase is expressed, which produces the mRNA of the target gene.

For expression *E. coli* BL21 (DE3) Gold was used in 250 mL terrific broth (TB) medium with kanamycin (50 $mg \cdot L^{-1}$) and 0.1 mM IPTG. 2 mL overnight culture were used as inoculum and the expression cultures were incubated for 24 h at 22 °C with 200 rpm.

Alternatively, at the beginning of the project, the protocol by Picart et al.[30] was used for enzyme expression. In this protocol the cells were grown to an OD_{600} of 0.6 at 37 °C and then the temperature was reduced to 22 °C and 0.1 mM IPTG were added.

2.3.5.4 Enzyme purification

Immobilized ion affinity chromatography (IMAC) was used for enzyme purification. The target protein is expressed with a poly-histidine tag, which binds over a chelate effect to immobilized metal ions. A common strategy is the usage of Ni^{2+} ions immobilized over NTA (nitrilotriacetic acid) groups on sepharose beads. After binding of the target enzyme to the metal ions, other unbound cell parts are washed away. Subsequently, the target protein is eluted with imidazole or by pH decrease.

Purification using ÄKTA systems

For the purification of enzymes pellets of 250 mL *E. coli* culture were suspended in 20 mL binding buffer (20 mM potassium phosphate buffer, pH 7.4, 500 mM NaCl, 20 mM imidazole), 1 tablet protease inhibitor cocktail (Pierce Protease Inhibitor Tablets, Thermo Fisher Scientific, Waltham, USA-MA) was added, and the cells lysed by sonication (4 min active pulsing, 2 sec pulse, 5 sec break, 60% amplitude). The cell debris was removed by centrifugation (18,000×*g*, 20 min, 4 °C) and filtration (0.45 µm pore size). The cell lysate was applied subsequently to a 5 mL Histrapp-FF column (GE Healthcare - Solingen, Germany) using an ÄKTA Start system (GE Healthcare - Solingen, Germany). The sample application was performed with a flowrate of 2 $mL \cdot min^{-1}$. Unbound proteins were washed out with 10 column volumes (CV) binding buffer, and 5 CV 2.5% elution buffer (20 mM potassium phosphate buffer, pH 7.4, 500 mM NaCl, 500 mM imidazole) with 5 $mL \cdot min^{-1}$ flowrate. The target protein was step-eluted with 70% elution buffer and a flowrate of 1 $mL \cdot min^{-1}$ (fraction size 1 mL). The collected protein fractions

were analyzed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE, section 2.3.8.3). Fractions containing the target protein were combined, concentrated (Vivaspin, Sartorius - Göttingen, Germany, MWCO 10 kDa), desalted (PD-10 Column, GE Healthcare - Solingen, Germany) with 20 mm Tris, 20% Glycerol pH 7.5 as eluent, and quantified using a Nanophotometer NP80 (Implen - München, Germany).

In the case of LigF215 the protocol was modified to a three buffer system. For the sonication, sample application, and washing (10 CV) the normal binding buffer (20 mM potassium phosphate buffer, pH 7.4, 500 mM NaCl, 20 mM imidazole) was used. Then, the buffer was changed to buffer B (20 mM Tris, 100 mM NaCl, 2 mM GSH, 20% Glycerol pH 8) in a step over 5 CV (flow rate 5 mL·min^{-1}). For further washing 5 CV 2.5% buffer C (20 mM Tris, 100 mM NaCl, 2 mM GSH, 20% Glycerol, 500 mM Imidazole pH 8) in buffer B with 5 mL·min^{-1} flowrate was used. The protein was eluted with a step to 25% buffer C with a flowrate of 1 mL·min^{-1} in 1 mL fractions. Before the elution, 1 mL of buffer B were added to the 2 mL reaction tubes to dilute the imidazole directly.

Alternatively, the protocol by Picart et al.[30] was used for enzyme purification at the beginning of the project. In this protocol the elution was performed using a gradient. The gradient was used after the washing with binding buffer from 0 to 70% elution buffer in 20 CV with 2 mL·min^{-1} flowrate and 2 ml fraction size.

Purifications using gravity columns

Gravity columns were used for the purification of smaller amounts of enzyme. The pellet of 100 mL *E. coli* culture was resuspended, lysed, and clarified as described before. Ni Sepharose™ 6 Fast Flow (GE-Healthcare - Solingen, Germany) with a column volume of 2 mL was used as column resin. After application of the clarified cell lysate, the column was washed with 10 mL binding buffer and 5 mL 2.5% elution buffer. The protein was eluted with 70% elution buffer in 1 mL fractions. The protein content of the samples was visualized with Bradford solution. The protein was subsequently desalted and stored as described before.

2.3.6 Enzyme characterization

2.3.6.1 Activity assay

The activity of an enzyme is an important characteristic for the comparison of enzymes. The activity describes the reaction rate at the reaction start, since at time point zero no back reaction and saturation effect need to be considered.

Do determine the activity of the β-etherases and glutathione lyases, 1 mL bioconversion reactions were performed and analyzed by HPLC (section 2.3.8.1).

β-etherases

The reaction for the determination of β-etherases activity contained 100 mM glycine/NaOH buffer pH 9, 1 mM GSH, 0.4 mM racemic 2,6-MP-VG (dissolved in 40% DMSO/H_2O, DMSO end concentration 10%) and 5 µg·mL^{-1} enzyme at 25 °C and 800 rpm. The components and volume composition of the reaction is equivalent to the β-etherases reaction described in Table 2.19 (exception enzyme concentration). Samples for HPLC analysis were taken after 1, 3, 5, 7, 9, 11 min. The sample was pipetted in sulphuric acid (5% end concentration) to stop the reaction and centrifuged (13,000×*g*, 10 min) to remove the enzyme.

Glutathione lyases

The substrate for the glutathione lyase reaction is synthesized in a β-etherase reaction with the enzymes LigE and LigF-NA to generate racemic GS-VG. To test the activity of glutathione lyases towards one enantiomer of GS-VG only type of β-etherases was used in the β-etherases reaction (LigE or LigF-NA). The composition of both reactions is shown in Table 2.19.

Table 2.19 Components of the glutathione lyase activity assay and the β-etherases substrate reaction.

Component	Composition	Volume %	End concentration
β-etherase reaction			
Buffer	200 mM glycine/NaOH pH 9 20 mM GSH	50%	100 mM glycine/NaOH pH 9 10 mM GSH
Substrate	32 mM 2,6-MP-VG 40% DMSO	25%	8 mM 2,6-MP-VG 10% DMSO
β-etherase	100 µg·mL^{-1} per enzyme	25%	25 µg·mL^{-1} per enzyme

Component	Composition	Volume %	End concentration
Glutathione lyase reaction			
β-etherases reaction	100 mM glycine/NaOH pH 9 10 mM GSH 8 mM GS-VG 10% DMSO	50%	50 mM glycine/NaOH pH 9 5 mM GSH 4 mM GS-VG 5% DMSO
Buffer	200 mM glycine/NaOH pH 9 20 mM GSH	25%	50 mM glycine/NaOH pH 9 5 mM GSH
Glutathione lyase	20 µg·mL^{-1}	25%	5 µg·mL^{-1}

The reaction volume for the glutathione lyase activity assay was 1 mL and HPLC samples were taken after 1, 4, 7, 10, 13, 16 min. The sample was pipetted in sulphuric acid (5% end concentration) to stop the reaction and centrifuged (13,000×*g*, 10 min) to remove the enzyme.

2.3.6.2 Selectivity assay

E-Value

The E-value describes how selectively an enzyme uses one enantiomer of a racemic substrate mixture. Low E-values indicates low selectivity. An E-value of 1 describes a complete unselective reaction. For useful enantioselective reactions E-values above 15 are required. E-values of 200 or higher are considered to describe perfect selectivity. Since the analytical error increases exponentially with increasing E-values, due to the logarithmic formula, E-values higher than 200 are not used.[101]

The selectivity of the β-etherases was determined by using chiral HPLC analysis with a Chiralcel OD-RH column (L 150 mm, ID x, particle 5 µm, Daicel - Raunheim, Germany). For the determination of the selectivity, the biocatalytic reactions as described in section 2.3.6.1 (β-etherases activity assay) were performed. The only differences were that the substrate was dissolved this time in isopropanol (5% end concentration) and that just one sample after 4 h was taken. For the analysis an isocratic mixture of water and acetonitrile (70/30, V:V) was used as mobile phase with a flowrate of 1 mL·min^{-1}. The selectivity was calculated based on the decrease of the two 2,6-MP-VG enantiomers according to Chen et al. with following formula.[102]

$$E = \frac{\ln[(1-c)(1-e.e._s)]}{\ln[(1-c)(1+e.e._s)]}$$

R/S ratio

The selectivity for the glutathione lyases could not be calculated directly, since GS-(*R*)-VG and GS-(*S*)-VG could not be separated by HPLC. Therefore, the activity for both substrates was determined as described in section 2.3.6.1 and the ratio of both was calculated.

2.3.6.3 pH and temperature optimum

Enzymatic activity is both pH- and temperature-dependent. To determine pH and temperature optimum, the activity of the enzyme is measured at various pH and temperatures. The trend of the reaction rate depending on pH and temperature offers information about the optimal conditions for catalysis and the stability of the enzyme.

pH optimum

To measure the pH optimum of the β-etherases a fluorogenic assay with the substrate MU-VG was used. The reactions were performed in 96 well microtiter plates (flat, Sarstedt - Nümbrecht, Germany) and measured online in CLARIOstar 96 well microplate reader (BMG Labtech - Ortenberg, Germany). The reaction volume was 200 µL containing 100 mM buffer, 1 mM GSH, 0.1 mM racemic MU-VG (dissolved in 40% DMSO/H_2O, DMSO end concentration 10%). The reaction was performed at 25 °C. The amount of enzyme and the different buffers are given in Table 2.20 and Table 2.21. The fluorescence of the produced 4-methylumbiliferone was measured at 450 nm with an excitation wavelength of 360 nm. The first 5 min of the reaction were used to calculate the slope by linear regression. Relative activities were calculated by setting the highest slope of each enzyme to 100%. To exclude buffer effects a mixture of Tris and glycine buffer (100 mM each) was tested as well for pH 8.5, 9 and 9.5.

Table 2.20 Buffers used for pH optimum assay.

pH	Buffer	pH	Buffer
5	acetate	9	glycine
6	phosphate	9.5	glycine
7	phosphate	10	glycine
8	Tris	11	carbonate

Table 2.21 Enzyme amounts used for pH optimum assay.

Enzyme	Concentration [µg·mL^{-1}]	Enzyme	Concentration [µg·mL^{-1}]
LigE	300	LigF-NA	25
LigE179	300	LigF008	25
LigE283	300	LigF215	400
LigE491	300	LigF729	25
LigE760	400	LigF755	25
LigE889	300	LigF921	25
LigE915	300	LigF935	25
		LigF965	25

Temperature optimum

The optimal reaction temperature of the β-etherase reactions were determined using an absorbance-based assay with the substrate VN-VG. The reaction was performed in micro cuvettes (Brand - Wertheim, Germany) and measured online in Cary60 UV-vis spectrometer (Agilent Technologies - Waldbronn, Germany). The reaction volume was 400 µL containing 100 mM glycine/NaOH pH 9, 1 mM GSH and 0.5 mM racemic VN-VG (dissolved in 40% DMSO/H_2O, DMSO end concentration 10%). The enzyme concentrations used for the absorbance assay are listed in Table 2.22. The reaction progress was monitored by absorbance at 360 nm. The temperature was increased from 10 to 50 °C in 5 °C intervals. Before the start of the reaction, all reaction components were incubated separately for 5 min at the measuring temperature. The reaction was measured about 7.5 min and the slope of the first 2 min was taken. To determine the relative activity the highest slope of each enzyme was set to 100%.

Table 2.22 Enzyme concentrations used for the absorbance assay for the determination of the temperature optimum.

Enzyme	Concentration [µg·mL^{-1}]	Enzyme	Concentration [µg·mL^{-1}]
LigE	4	LigF-NA	4
LigE179	8	LigF008	1
LigE283	4	LigF215	10
LigE491	4	LigF729	4
LigE760	50	LigF755	4
LigE889	4	LigF921	4
LigE915	4	LigF935	1
		LigF965	4

2.3.6.4 Enzyme melting point (thermofluor assay)

The thermofluor assay (also called thermoshift assay) is used to determine the heat denaturation point (melting temperature T_m) of an enzyme. The assay uses a fluorescence dye, which is quenched in water. During the heating of the sample the enzyme gets unfolded, the hydrophobic enzyme core is released and the dye binds to these hydrophobic amino acids. Through this interaction the dye can emit fluorescence, which is measured.

The thermofluor assay was performed in a qPCR machine (CFX96 real time system c1000 thermal cycler, Bio-Rad Laboratories - Feldkirchen, Germany), with 5 times Sypro Orange (Invitrogen/Thermo Fisher Scientific - Waltham, USA-MA; 5000x concentrated) and 0.2 mg·mL^{-1} Protein end concentration. The sample had a total volume of 50 µL with 40 µl of the tested buffers, 5 µL protein solution and 5 µL Sypro Orange (1/100 diluted). For the determination of the melting point 20% glycerol in 20 mM Tris pH 7.5 was used as buffer, which was also used for storage and dilution of the proteins. The qPCR machine performed a temperature gradient from 10 to 90 °C, while measuring the fluorescence with an excitation wavelength of 490 nm and an emission wavelength of 575 nm.

2.3.6.5 Micro scale electrophoresis MST

Micro scale electrophoresis (MST) is a method to analyse and quantify ligand-receptor binding. In this method the temperature-induced change in fluorescence of a fluorogenic dye, which is bound to the receptor, is used to measure the ligand binding. The ligand binding has an influence on the change in fluorescence and, therefore, can be determined.

Purified LigF-NA was adjusted to a concentration of 200 mM using Tris buffer (20 mM Tris, 100 mM NaCl, pH 7.5). The enzyme solution was mixed 1/1 with RED-tris-NTA dye (Nanotemper - München, Germany; diluted with PBS-T: 137 mM NaCl, 2.5 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4, 0.05% Tween 20; end concentration 100 mM) and incubated on ice for 30 min. 10 µL of inhibitor **11** (section 2.3.1.4) were added to 10 µL of the enzyme-dye-mixture. For this 2 mM **11** were solved in 10% DMSO and a 1/1 dilution-series performed (16 different concentrations). The different mixtures were centrifuged and filled in capillaries (Nanotemper - München, Germany). The measurement was performed with an MST device Monolith NT.115 from Nanotemper (München, Germany) with 60% LED power and 40% power of the IR-laser.

2.3.7 Protein engineering

2.3.7.1 Generation of SSM-mutant library

The insertion of site-directed saturation mutations was performed with PCR reactions using denatured primers (section 2.3.4.5). After transformation in *E. coli* XL1-Blue cells (section 2.3.4.4), the mutant library was isolated as a plasmid (section 2.3.4.6). The library was subsequently transformed in *E. coli* BL21 (DE3) Gold cells (section 2.3.4.4). Single colonies were picked form agar plates and transferred into a 96 microtiter plate containing 200 µL LB media with kanamycin (layout shown in Table 2.23). The bacteria cells were cultivated overnight (section 2.3.7.2). For storage a cryo plate (master plate) was produced (section 2.3.5.2) and sealed with SILVERSEAL SEALER Aluminium foil (Greiner Bio-One - Frickenhausen, Germany). The quality of the library was proven by sequencing (section 2.3.4.7) of the plasmid library and 3 to 5 single colonies.

A disadvantage of the direct transformation of the PCR product in *E. coli* BL21 (DE3) Gold cells is the low transformation efficiency compared to *E. coli* XL1-Blue cells. Therefore, the PCR product was first transformed in *E. coli* XL1-Blue, the mutant library isolated as plasmid, and subsequently transformed in *E. coli* BL21 (DE3) Gold to reach the needed number of colonies.

Table 2.23 Layout of master plate used for protein engineering. (WT = wild-type control, EV = empty vector control)

	1	2	3	4	5	6	7	8	9	10	11	12
A	**WT**	**EV/WT**	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
B	C11	C12	**WT**	**EV/WT**	C13	C14	C15	C16	C17	C18	C19	C20
C	C21	C22	C23	C24	C25	C26	C27	C28	C29	C30	C31	C32
D	C33	C34	C35	C36	**WT**	**EV/WT**	C37	C38	C39	C40	C41	C42
E	C43	C44	C45	C46	C47	C48	**WT**	**EV/WT**	C49	C50	C51	C52
F	C53	C54	C55	C56	C57	C58	C59	C60	C61	C62	C63	C64
G	C65	C66	C67	C68	C69	C70	C71	C72	C73	C74	C75	C76
H	C77	C78	C79	C80	C81	C82	C83	C84	C85	C86	**WT**	**EV/WT**

2.3.7.2 Overnight cultures for protein engineering

Overnight cultures of the mutant libraries were grown in microtiter plates. 200 µL of LB-media with kanamycin were either inoculated with transformed colonies of an agar plate, as described in section 2.3.7.1, or from a frozen master plate (cryo culture, section 2.3.5.2), respectively. The culture was cultivated over night at 37 °C and 850 rpm.

2.3.7.3 Expression in 96 well plates

The expression of the mutated enzymes in 96 microtiter plates with *E. coli* BL21 (DE3) Gold was performed in 180 µL LB media with kanamycin (50 $mg \cdot L^{-1}$) and 0.1 mM IPTG, which were inoculated with 20 µL overnight culture (section 2.3.7.2). The plate was incubated for 24 h at 30 °C with 850 rpm. During the incubation the culture was sealed with a breathable rayon film (VWR - Darmstadt, Germany; Cat. No 60941-084). The final OD_{600} was measured by diluting the bacterial cells 1/10 with MilliQ water. The cells were pelleted (4,000×*g*, 20 min, 4 °C) and stored at -20 °C.

2.3.7.4 Glutathione reductase assay

The bacteria colonies were resuspended and lysed in 200 µL B-PER (bacterial protein extraction reagent, Thermo Fisher Scientific - Waltham, USA-MA). The plate was shaken at room temperature with 800 rpm for 20 min. After centrifugation (4,000×*g*, 20 min, 4 °C) the supernatant was diluted 1/10 and used for the reaction (composition of the reaction in Table 2.24). The reaction was monitored in a CLARIOstar 96 well microplate reader (BMG Labtech - Ortenberg, Germany) at 360 nm over 30 min. To start the reaction, the substrate (β-etherase reaction mixture, Table 2.19) was added.

The relative activity was determined by linear regression of the reaction curve (1.5 to 15 min). The slope was normalized to the OD_{600}. Afterwards, for a better comparison the difference of this value to the mean value of the wild-type controls was calculated in percent (formula is shown below).

$$\left[\left(\frac{Slope/_{OD_{600}}}{Mean\left(WT\ controls/_{OD_{600}}\right)}\right)-1\right]\times 100$$

Table 2.24 Composition of glutathione reductase assay reaction.

Component	Composition	Volume	End concentration
Buffer	200 mM Tris pH 8.5 20 mM GSH	50 µL	50 mM Tris pH 8.5 5 mM GSH
NADPH and glutathione reductase	9.6 mM NADPH 1/100 glutathione reductase	25 µL	1 mM NADPH 0.25 µL glutathione reductase
Cell-free extract	Diluted 1/10	25 µL	
β-etherase reaction	100 mM Tris pH 8.5 10 mM GSH 8 mM GS-VG 10% DMSO	100µL	50 mM Tris pH 8.5 5 mM GSH 4 mM GS-VG 5% DMSO

2.3.7.5 Mutant validation

Single mutants, which showed high activity in the glutathione reductase assay, were validated by expressing the enzymes in 100 mL TB media, as described in section 2.3.5.3 followed by protein purification over gravity columns (section 2.3.5.4), and activity measured via the HPLC assay (section 2.3.6.1).

2.3.7.6 GSH detection assay

A kinetic assay, developed by Rahman et al., was tested for the quantification of GSSG.[103] In this assay, free GSH is derivatized with 2-venylpyridine. GSSG is reduced by the glutathione reductase to GSH. GSH then reacts with DTNB (5,5'-dithiobis-(2-nitrobenzoic acid)) and forms a GSH-adduct as well as 2-nitro-5-thiobenzoate (TNB), which is measured. The reaction speed of this reductive cleavage (measured by the TNB production) is dependent on the GSSG concentration and refers to the GSSG start concentration (see also section 3.4.3.1).

40 µL sample were pipetted in 40 µL H_2SO_4 / 2-venylpyridine mixture (2% H_2SO_4, 270 mM 2-venylpyridine, 17% DMSO). After 5 min of incubation 40 µL 0.73 M triethanolamine were added, which resulted in a pH around 7. The mixture was sealed (SILVERSEAL SEALER Aluminium foil, Greiner Bio-One - Frickenhausen, Germany) and incubated for 1h. Next, 45 µL glutathione reductase and DTNB were added (0.31 µL glutathione reductase → 0.007µL glutathione reductase per reaction, 2.66 µL DNTB → 0.6 mM end concentration, in 500 mM KPi pH 7.5). To start the reaction 35 µL 3.43 mM NADPH in 500 mM KPi pH 7.5 were added. The absorbance was measured online at 412 nm for 30 min.

2.3.8 Analytics

2.3.8.1 HPLC analytics

High performance liquid chromatography (HPLC) is a technique for the determination of concentration and purity of solubilized compounds. The sample is pushed by the mobile phase through a stationary phase. The interaction of the sample with the stationary phase hinders the migration of the sample in the column. The strength of this interaction depends on the polarity of the analyzed compounds, which leads to the separation of the different compounds. The used combination of the polar mobile and unpolar stationary phase is called reverse phase chromatography. In HPLC columns the stationary phase consists of very small particles and by applying high-pressure a good separation in short time is achieved.

The samples were analyzed with Nerxera XR20 system (Shimadzu - Duisburg, Germany) with a Nucleosil C18 HPLC column (L 250 mm, ID 5 mm, particle 5 µm, Macherey Nagel - Düren, Germany). As a mobile phase, a isocratic mixture of water, acetonitrile and TFA (49.95/50/0.05, V:V:V) with a flowrate of 1 $mL \cdot min^{-1}$ was used. The concentration of 2,6-MP-VG, 2,6-dimethoxyphenol, GS-VG and VG were determined based on calibration curves, detected at 280 nm. For the calibration curves chemical synthesized 2,6-MP-VG, biocatalytical produced GS-VG and VG and commercial 2,6-dimethoxyphenol were used.

The chiral HPLC analysis for the determination of the E-value of β-etherases is described in section 2.3.6.2. Typical HPLC chromatograms and the corresponding calibration curves for all HPLC measurements are shown in section 6.3.

2.3.8.2 External analysis

LC-MS analyses were performed by Ulrike Beutling in cooperation with the Department of "Chemical Biology" at the Helmholtz Center for Infection Research in Braunschweig in the group of Prof. Mark Brönstrup. The group of Prof. Martin Bröring in the Institute of "Inorganic and Analytic Chemistry" of the TU Braunschweig performed the ESI-MS analysis. HRMS-ESI-MS and NMR analysis were performed in the central facilities of the TU Braunschweig.

2.3.8.3 Electrophoresis

For the analysis of DNA and protein samples agarose gel electrophoresis or SDS-PAGE (SDS-polyacrylamide gel electrophoresis) were performed, respectively.

Agarose gel electrophoresis

Agarose gel electrophoresis separates DNA fragments by their size. DNA molecules are negatively charged due to their phosphoric acid diester backbone and migrate in the electric field to the anode. The size separation of the charged DNA molecules is provided by an agarose gel network, which hinders bigger molecules more than smaller molecules in their migration. For the analysis, agarose in the concentration of 1% was used. The agarose was filled up with 1x TAE buffer and heated until it was completely dissolved. The hot agarose was used to form the gel. For staining Midori Green Direct (Nippon genetics - Düren, Germany) was added to the 6x sample buffer (1:10 ratio) and the buffer was further used as 5x buffer. The loaded samples were separated by applying 150 V to the gel. The gel was visualized via GelPic LED Box.

SDS-PAGE

SDS-PAGE was performed for the analysis of protein samples. A gel of acrylamide and bis-acrylamide, crosslinked by tetramethylethylendiamine (TEMED) and ammonium persulfate (APS), hinders the migration of protein molecules in the electric field dependent on their size. As result, the smaller molecules migrate faster than bigger molecules. Since the folding of proteins could influence the migration behaviour, the proteins are denatured by SDS and β-mercaptoethanol. SDS also covers the protein sample with a negative charge, which is important for the migration in the electric flied.

The samples were mixed with 4x SDS-buffer, heated for 10 min at 95 °C, and shortly centrifuged. The loaded samples were separated in the gel at 200 V. The gel was stained with

coomassie staining solution. Afterwards the gel was destained with destaining solution and water. Table 2.25 shows the recipe for the SDS-gel preparation.

Table 2.25 Components for the preparation of 4 SDS gels.

	Separating gel	**Stacking gel**
Tris pH 8.8	5.0 mL	-
Tris pH 6.8	-	2.5 mL
MilliQ	9.0 mL	6.0 mL
APS 10%	0.2 mL	0.1 mL
TEMED	0.02 mL	0.01 mL
Acrylamid 40%	6.0 mL	1.5 mL

2.3.9 Protein crystallization

X-ray crystallography is a technique to determine the structure of molecules. The X-ray beam is diffracted on the electron shells of atoms, which form the target molecule. The crystal works as an enhancer of the diffraction. The crystal-structure is the determined based on the diffraction pattern and phases. To solve the phase problem, either phases have to be measured by additional crystallization and X-ray experiments or the phases of known crystal structures are used. Using both information the electron density of the molecule can be calculated, which is used to build the protein structure.

The crystallization of LigG-TD was performed in cooperation with the work group “Structur und Function of Proteins” of Prof. Wulf Blankenfeld (HZI, Braunschweig) with the co-worker Christina Diederich. The affinity purification was performed using the gradient elution protocol in the workgroup of Prof. Anett Schallmey (section 2.3.5.4). The GPC and the crystallization were performed together with Christina Diederich in the HZI. 20 mM Tris/HCl, pH 7.5, 100 mM NaCl were used as solvent for the GPC. For the crystallization the JCSG Core I-IV (Qiagen - Hilden, Germany), MORPHEUS (Molecular Dimensions - Suffolk, UK), and the MIDAS (Molecular Dimensions - Suffolk, UK) crystallization screens were used. The drop contained 0.2 µL reservoir solution and 0.2 µL protein solution. The concentrations of the protein solutions were 20 $mg \cdot mL^{-1}$ and 10 $mg \cdot mL^{-1}$, respectively. Additionally, one protein solution of 10 $mg \cdot mL^{-1}$ was used containing 10 mM glutathione. The screens were stored at 20 °C. The measuring of the X-ray diffraction dataset and the build of the protein structure was performed by Christina Diederich.

3 Results

The results of this thesis are structured in different subprojects. As basis for all further studies, first the synthesis of lignin model compounds as well as the expression and purification of β-etherases and glutathione lyases was optimized (sections 3.1 and 3.2). The main work of this thesis consists in the identification of new lignin-degrading enzymes and the optimization of the glutathione lyase LigG-TD by protein engineering. The identification of new putative β-etherases in public databases (section 3.3) was performed by a combination of peptide pattern recognition and phylogenetic analyses. A representative subset of these enzymes was biochemically characterized, and a sequence-structure-function analysis was performed.

Regarding the optimization of LigG-TD by protein engineering (section 3.4), a high-throughput screening assay was developed, and the crystal structure of LigG-TD was solved. Based on this structure and with the help of the developed screening assay, the activity of LigG-TD was increased in one round of protein engineering. Additionally, also the family of glutathione lyases was phylogenetically analyzed and two Nu-class glutathione lyases were biochemically characterized (section 3.5).

Moreover, in further side projects a β-*S*-4 aryl thioether compound was tested as possible substrate for β-etherases (section 3.6) as well as the synthesis of a β-etherase inhibitor was attempted (section 3.7).

3.1 Optimization of lignin model substrate synthesis

Different protocols for the synthesis of various lignin model compounds have been published.[30,31] The protocol by Picart et al.[30] was herein used as starting point for further optimization. The synthesis consists of three steps (Figure 3.1): bromination of an acetophenone derivate, formation of the keto ether, and hydroxy methylation. Especially the bromination and hydroxy methylation reaction still offered a huge potential for improvement.

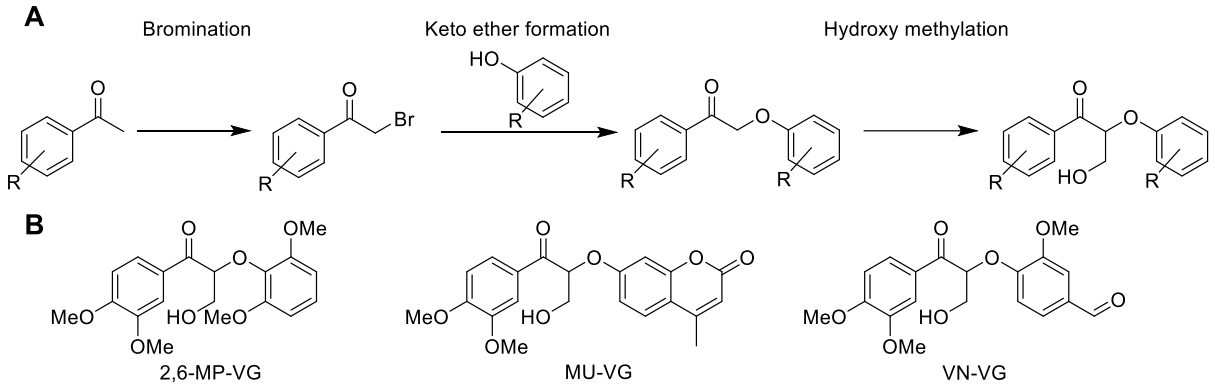


Figure 3.1 **A** General reaction scheme of lignin model substrate synthesis with the three reaction steps. **B** Structure of the three lignin model substrates investigated in this part of the thesis (abbreviations and full names of the lignin model substrate are listed in section 2.3.1).

The bromination of an acetophenone derivate depends on water free conditions and is the most elaborate step in the synthesis route. Therefore, it is very important to synthesize high amounts of brominated acetophenone in one reaction. To reach that goal the reaction volume was doubled and the substrate concentration was increased by 50%, resulting in a roughly three times higher amount of product per reaction without compromised isolated yield of 80%, comparted to the yields of Picart et al.[30]

The second step, formation of the keto ether, was performed according to Picart et al. with nearly quantitative yields.[30] Hence, no significant changes were made in this reaction step.

The last step in the model substrate synthesis is the hydroxy methylation of the keto ether. In most cases only moderate yields were obtained by Picart et al.[30] Furthermore, the solvent DMSO, which was used in the synthesis, is challenging during workup and double hydroxy methylations as well as other side reactions occurred. Therefore, a solvent and base screening was carried out to improve the hydroxy methylation reaction. Hence, reactions in 1 mL scale with each 10 mg substrate were performed using different solvents and bases analyzed by HPLC. Interestingly, the used solvent had a significant impact on the reaction outcome, while for the bases only carbonate was effective. In the synthesis of VN-VG each tested solvent yielded in a different pattern of product and side products. Whereas for example in DMSO no desired product formation was observed but only side products were formed, in isopropanol exclusively the desired product was obtained (Figure 3.2).

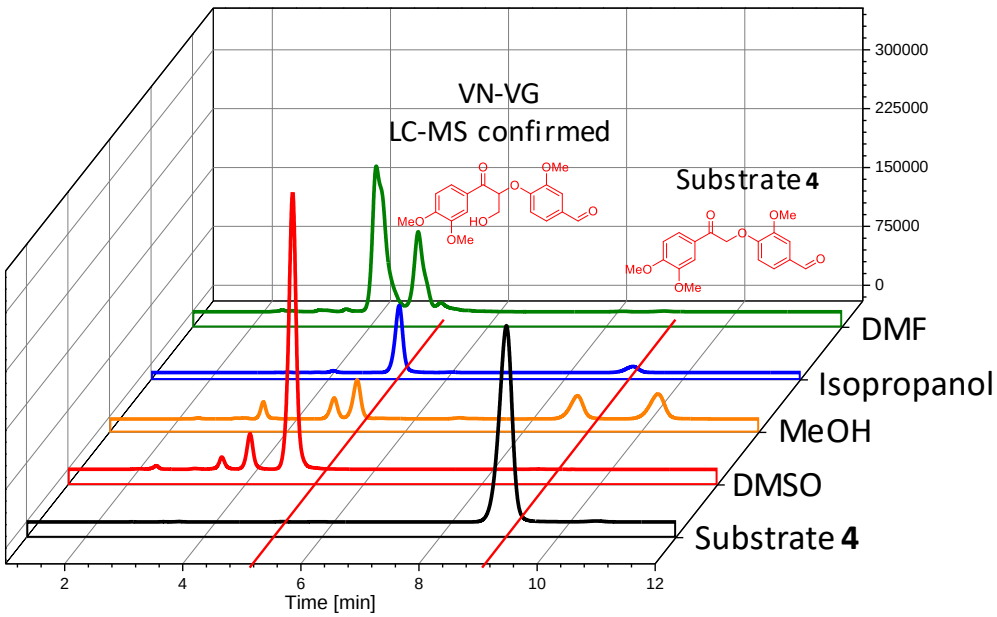
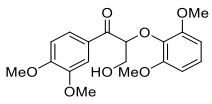
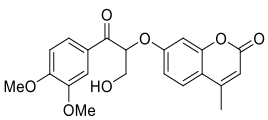
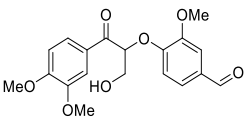
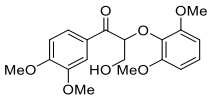
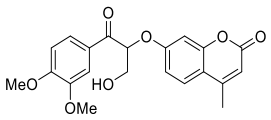
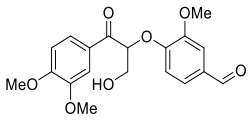


Figure 3.2 Solvent screening for the hydroxy methylation reaction of VN-VG synthesis. HPLC chromatograms of reactions performed in 1 mL scale using each 10 mg substrate **4** (black). The different solvents lead to different reaction products. In the case of methanol (orange) and isopropanol (blue), VN-VG is the main product of the reaction, whereas it is not formed or only present in traces using DMSO (red) or DMF (green) as solvents. The HPLC measurement was performed with an isocratic mixture of 40% ACN in water with 0.1% TFA as solvent using the same HPLC-system described in section 2.3.8.1.

Since no general working reaction conditions for the third step could be identified, the reaction conditions had to be optimized for every model compound individually. Table 3.1 shows a summary of all performed solvent screenings for the different hydroxy methylation reactions.

Table 3.1 Solvent screening results for the hydroxy methylation reactions in the synthesis of three lignin model compounds. The table lists the percentage of formed product; the percentage of formed side products (SP) is given in brackets. The best conditions are highlighted in bold. Reactions were performed in 1 mL using each 10 mg substrate and were analyzed by HPLC after 1 day of reaction.

	 2,6-MP-VG	 MU-VG	 VN-VG
Base	DIPEA	Na_2CO_3	Na_2CO_3	Na_2CO_3
Acetone	1% (35% SP)	2% (44% SP)	-	n.d.
ACN	4	5	-	27
DCM	1% (7% SP)	2% (14% SP)	-	n.d.
DMF	n.d.	5%	-	4% (96% SP)
DMSO	1%	19%	**63% (37% SP)**	0% (100% SP)
EtOAc	1%	2%	-	n.d.
EtOH	n.d.	n.d.	n.d.	65% (35% SP)
Isopropanol	6% (3% SP)	14% (3% SP)	15% (5% SP)	**87%**

	 2,6-MP-VG		 MU-VG	 VN-VG
Methanol	5	**100%**	17% (17% SP)	25% (51% SP)
MTBE	0%, (4% SP)	1%	-	11% (60% SP)
THF	1% (2% SP)	1%	-	24%

n.d. = no reaction detected; SP = side product

The optimized reaction conditions for the hydroxy methylation reaction were also used for preparative syntheses of 2,6-MP-VG and VN-VG resulting in 84% and 75% isolated yield, respectively. In summary, an efficient synthesis of lignin model compounds was possible after optimization of the reaction conditions.

3.2 Optimization of expression and purification of β-etherases and glutathione lyases

Expression and purification of β-etherases and glutathione lyases were performed using established protocols and vector systems provided by Picart et al.[30] (section 2.3.5). The enzymes were expressed in *E. coli* BL21 (DE3) Gold using pET28a vectors harbouring codon optimized genes for the expression in *E. coli*.

Following the protocol, the expression culture was first incubated at 37 °C until an OD_{600} of 0.6. Then, the expression was induced with 0.1 mM IPTG and the temperature was reduced to 20 °C for 20 h of incubation. The overexpression of β-etherases and glutathione lyases was successful and high amounts of soluble protein could be obtained. After IMAC purification, using Ni-NTA columns and an imidazole gradient, moderate to high yields of 6 to 100 mg protein per litre culture were obtained in high purity. In the case of LigF-NA proteolytic digestion was observed, which could be prevented using protease inhibitor tablets, while the addition of PMSF (phenylmethylsulfonyl fluoride) did not have any effect. Thus, protease inhibitor tablets were used for all further purifications.

To reduce the work expense and time for protein production, the expression protocol was modified. IPTG was added directly to the expression medium, which was inoculated with 2% overnight culture. Then, the culture was incubated at 22 °C for 24 h. This protocol did not only avoid regular measurements of optical density, it also resulted in increased protein yields two to four times in the case of LigE and LigF-NA (Table 3.2).

Regarding the IMAC purification, using an imidazole gradient for protein elution, as in the protocol by Picart et al.[30], resulted in broad elution peaks and a high protein elution volume (30 to 60 mL), making subsequent concentration and desalting steps time consuming. In contrast, a step-wise elution using 356 mM imidazole (70% elution buffer) instead of a gradient yield a concentrated enzyme solution in a volume of about 10 mL without compromised enzyme purity. In Figure 3.3, representative purification chromatograms and SDS-PAGE analyses for both protocols are shown. Table 3.2 summarizes the purified protein yield of different β-etherases and glutathione lyases using either a gradient or step-wise elution.

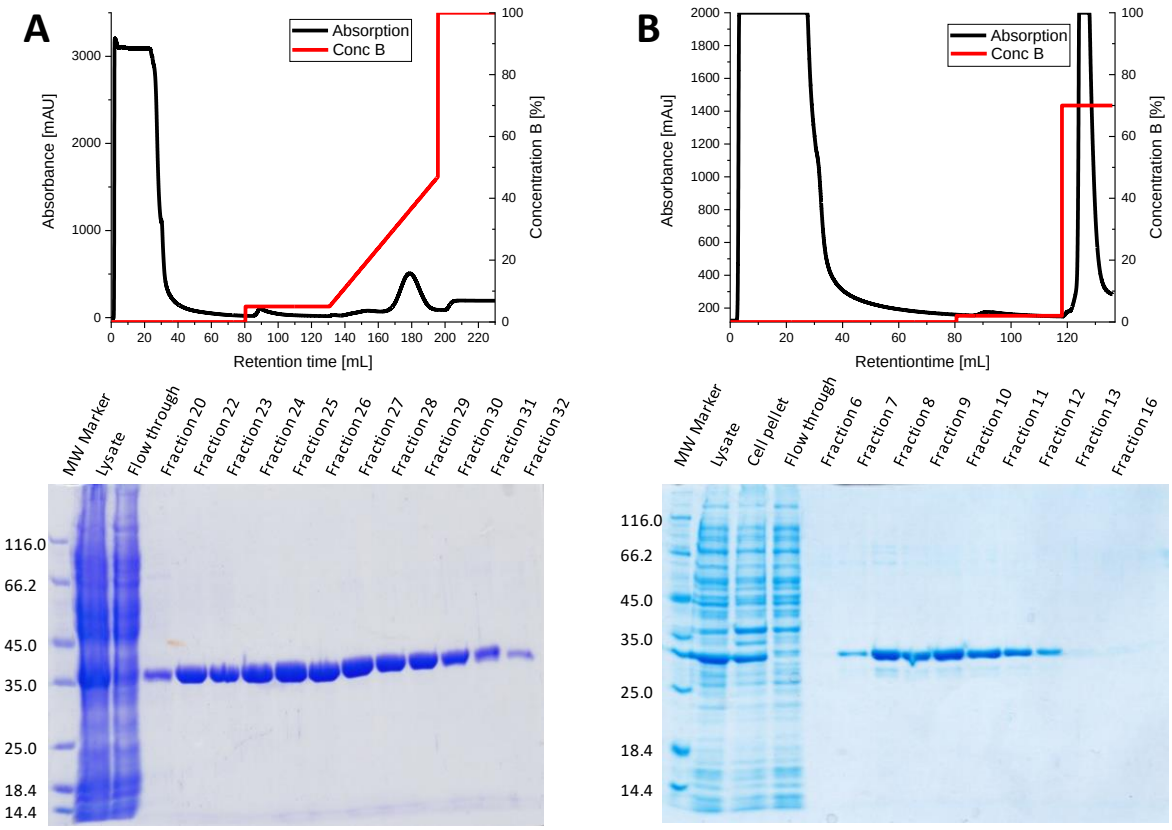


Figure 3.3 Chromatograms of affinity purifications of two β-etherases using HisTrap FF columns and an ÄKTA protein chromatography system (GE Healthcare - Solingen, Germany), as well as corresponding SDS-PAGE analyses (sections 2.3.5.4 and 2.3.8.3). **A** Purification of LigF-NA with gradient elution, **B** purification of LigE889 with step-wise elution. In both case a clear protein elution peak was observed, and the purity of both enzymes was confirmed by SDS-PAGE. The SDS-PAGE gels were stained with Coomassie Brilliant Blue and the Pierce™ Unstained Protein Molecular Weight Marker (Thermo Fisher Scientific - Waltham, USA-MA) was used as size as reference.

Table 3.2 Protein yields after affinity purifications of β-etherases and glutathione lyases.

Enzyme	Protein yield	Enzyme	Protein yield
LigE	62 mg·L^{-1}	LigF-NS	125 mg·L^{-1}
LigE*	114 mg·L^{-1}	LigF-NA	51 mg·L^{-1}
LigE-NS	27 mg·L^{-1}	LigF-NA*	227 mg·L^{-1}
LigE-NA	82 mg·L^{-1}	LigG	56 mg·L^{-1}
LigP	17 mg·L^{-1}	LigG-NS	6.1 mg·L^{-1}
LigF	249 mg·L^{-1}	LigG-TD	34 mg·L^{-1}

* purification with imidazole step

The activity of all enzymes was confirmed either by fluorescence or HPLC based assays using MU-VG and 2,6-MP-VG as substrates, respectively. The purified enzymes were used in the case of LigE, LigF-NA, LigG, and LigG-TD as reference enzyme for the characterization of new β-etherases (section 3.3) as well as for new glutathione lyases (section 3.5). Furthermore, all these enzymes were tested in the conversion of thioether model compounds (section 3.6).

3.3 Identification of new β-etherases

The number of known β-etherases is very limited. At the beginning of this project only five LigE-type (LigE, LigE-NA, LigE-NS, LigP, GST5), five LigF-type enzymes (LigF, LigF-NA, LigF-NS, NaLigF-1, GST4), and NaLigF-2 were known. To increase the knowledge about this enzyme group and to find more members with interesting or better catalytic properties, the goal was to screen public sequence data bases for novel sequences.

3.3.1 Identification of new putative β-etherases by PPR analysis

To identity new putative β-etherases, the PPR (peptide pattern recognition) algorithm was applied. The PPR algorithm clusters proteins based on the conservation of small peptide sequences instead of sequence alignments that are used in BLASTP searches.

Two separate databases for the PPR analysis were generated through BLASTP searches in the nr database of GenBank (release 226) using all known LigE-, LigF-, and NaLigF-2-type sequences as queries (BLASTP size 1,000 sequences). In case of the LigE-type database BLASTP resulted in 1,171 unique proteins with 27,967 hexamer peptides in total. PPR analysis clustered these into 11 groups of which one group comprised all known LigE-type enzymes together with 45 new sequences. For LigF-type enzymes the database consisted of 1,956

individual enzymes with 64,584 hexamer peptides in total. Here, the PPR algorithm created 54 groups of which one group contained all known LigF-type enzymes and 51 other enzyme sequences. NaLigF-2 was not grouped together with the LigF-type enzymes.

Most of the new putative enzymes have their origin in alphaproteobacteria of the order *Sphingomonadales*. In addition to *Sphingobium* or *Novosphingobium* species, now also *Erythrobacter* and *Altererythrobacter* species were found as host organisms. Interestingly, putative β-etherase-encoding sequences WP_104830666 (LigE-type) and WP_104831261 (LigF-type) are derived from the genome sequence of *Marinicaulis flavus*, a marine alphaproteobacterium of the family *Parvularculaceae*. Additionally, the host organism of LigF215 (GenBank accession number OGT78215) is classified as gammaproteobacterium.

In 31 bacterial strains sequences of both groups, LigE- and LigF-type β-etherases, were found, which is important since (*R*)- und (*S*)-selective β-etherases are required by bacteria for effective lignin utilization. Examples for bacterial strains with putative LigE- and LigF-type enzymes are *Marinicaulis flavus*, *Altererythrobacter* sp. 66-12, *Altererythrobacter* sp. Root672 and *Novosphingobium* sp. SCN 63-17. The GenBank accession number of all LigE- and LigF-type enzymes, group by PPR, as well as all genes used for the creation of the used database are listed in section 6.4.

Due to the large number of new putative β-etherases, 96 in total, it was not possible to order synthetic genes for all and to test all individual enzymes for β-etherase activity. Instead, a representative set of sequences was chosen to evaluate the suitability of the used method for identification of novel β-etherases. For further analysis and to aid in the selection process for representative enzymes, all sequences were aligned using webPRANK as a basis for maximum likelihood tree building by the IQ-TREE webserver (Figure 3.4). Based on this phylogenetic analysis a set of six putative LigE enzymes (LigE179, LigE283 LigE491, LigE760, LigE889, and LigE915) and seven putative LigF enzymes (LigF008, LigF215, LigF729, LigF755, LigF921, LigF935, and LigF965) were chosen for further analysis with LigE as well as LigF-NA used as reference enzymes (Tab. 1).

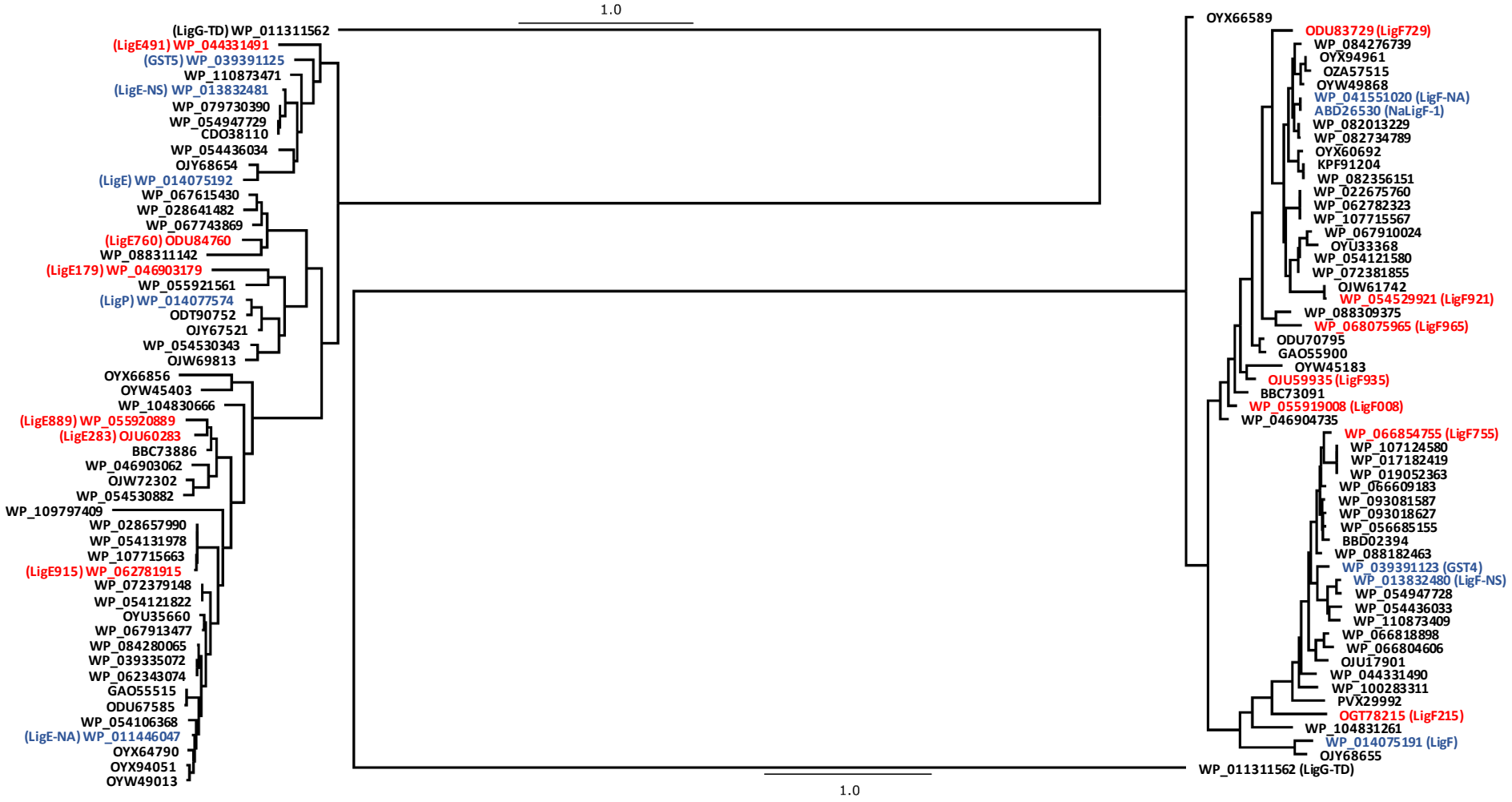


Figure 3.4 Maximum likelihood trees of putatively novel as well as known β-etherases using glutathione lyase LigG-TD from *Thiobacillus denitrificans* as outgroup. The left tree includes all LigE-type enzymes, whereas the right tree covers all LigF-type enzymes. Previously known β-etherases are highlighted in blue, while those enzymes that were selected for further characterization within this study are marked in red. Alignment and the creation of the phylogenetic tree was performed with the webtools webPRANK and IQTREE (section 2.3.3.3).[95,96] Respective GenBank accession numbers of the sequences included in both phylogenetic trees are listed in section 6.4

Table 3.3 List of putatively novel β-etherases chosen for characterization as well as reference enzymes (*).

Name	NCBI number	Organism
LigE-type enzymes		
LigE*	WP_014075192	*Sphingobium sp. SYK-6*
LigE179	WP_046903179	*Altererythrobacter atlanticus*
LigE283	OJU60283	*Altererythrobacter* sp. 66-12
LigE491	WP_044331491	*Sphingomonas hengshuiensis*
LigE760	ODU84760	*Novosphingobium* sp. SCN 63-17
LigE889	WP_055920889	*Altererythrobacter* sp. Root672
LigE915	WP_062781915	*Novosphingobium capsulatum*
LigF-type enzymes		
LigF-NA*	WP_041551020	*Novosphingobiu. aromaticivorans* DSM 12444
LigF008	WP_055919008	*Altererythrobacter* sp. Root672
LigF215	OGT78215	Gammaproteobateria
LigF729	ODU83729	*Novosphingobium* sp. SCN 63-17
LigF755	WP_066854755	*Sphingobium* sp. TCM1
LigF921	WP_054529921	*Erythrobacter* sp. SG61-1L
LigF935	OJU59935	*Altererythrobacter* sp. 66-12
LigF965	WP_068075965	*Novosphingobium lentum*

Next to the increased number of putative β-etherases, also the sequence diversity is increased. In the test set, the lowest sequence identities among LigE- and LigF-type enzymes are 53% and 54%, respectively. Sequence identities for the whole set of putative novel β-etherases go as low as 49% in both cases.

3.3.2 Cloning, expression, and purification of putative β-etherases

The selected putative β-etherases were ordered as synthetic genes with codon optimization for *E. coli* and cloned into expression vector pET28a. For this the vector as well as the synthetic genes were digested with NdeI, and HindIII and ligated using T4-DNA ligase (sections 2.3.4.1, 2.3.4.2, and 2.3.4.3). Colony PCR was used to select clones with successfully ligated vector and insert (Figure 3.5). Successful cloning was further confirmed by sequencing of positive clones by colony PCR (section 2.3.4.5).

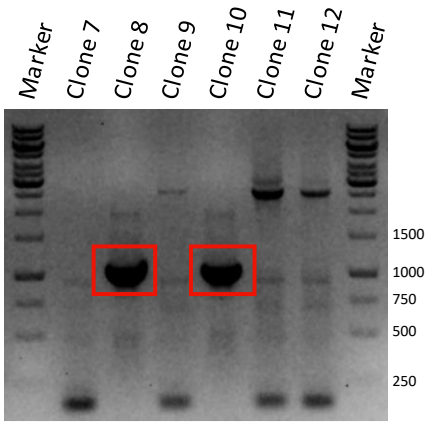


Figure 3.5 Colony PCR for cloning of the LigE283 gene into pET28a. The PCR product was separated in a 1% agarose gel in an electric field and stained with Midori Green (section 2.3.8.3). Colonies 8 and 10 show bands of the expected size (approximately 1000 bp).

Recombinant expression of the β-etherases in *E. coli* Bl21 (DE3) Gold and their purification based on IMAC were performed according to the improved protocols (expression at 22 °C and elution using an imidazole step, sections 2.3.5.3 and 2.3.5.4). The ÄKTA chromatograms for protein purification as well as SDS-PAGE gels were comparable to the results shown in section 3.2 and are therefore not displayed. The yield of purified enzymes ranged between 72 to 497 $mg \cdot L^{-1}$ culture except for LigF965, for which only 31 $mg \cdot L^{-1}$ were obtained (Table 3.4). Just in the case of LigF215, the general purification protocol had to be modified due to the enzyme's low stability indicted by enzyme precipitation during purification. To increase the stability of LigF215 during purification, the influence of imidazole, GSH, and glycerol on the stability of LigF215 was investigated using the thermofluor assay (Figure 3.6, section 2.3.6.5).

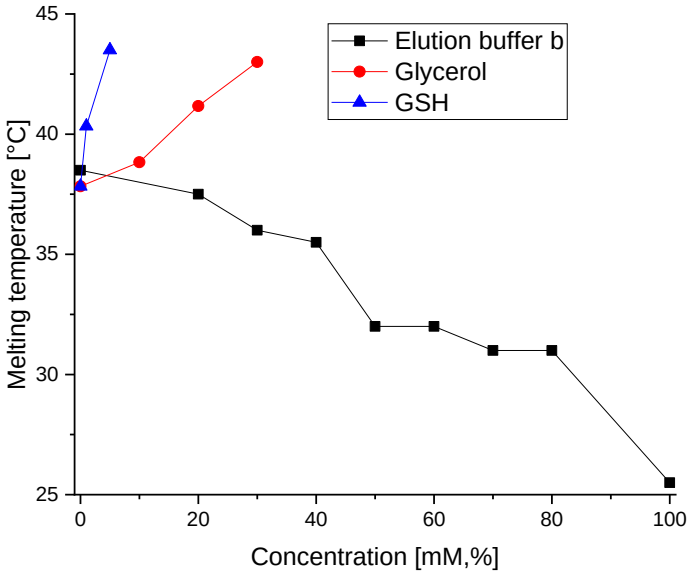


Figure 3.6 Influence of GSH, glycerol, and imidazole on the stability of LigF215. LigF215 stability was investigated using the thermofluor assay (section 2.3.6.5). Different concentrations of GSH (blue; 0, 1, and 5 mM) and glycerol (red; 0, 10, 20, and 30%) were tested in 20 mM Tris buffer pH 7.5. The elution buffer B (20 mM potassium phosphate buffer, pH 7.4, 500 mM NaCl, 500 mM imidazole) was tested in a concentration range from 0 to 100% (20 to 500 mM imidazole; black).

The thermofluor assay results show that GSH and glycerol have stabilizing effects on LigF215, whereas imidazole (elution buffer b) leads to a drastic destabilization. Hence, during IMAC purification 1 mM GSH and 20% glycerol were added to the buffers to stabilize LigF215. Furthermore, the maximal imidazole concentration for the elution of LigF215 was reduced to 125 mM. Using these conditions, no precipitation occurred and LigF215 was obtained in a stable and soluble form.

Table 3.4 Protein yields per litre culture of putative β-etherases after affinity purifications using improved protocols for expression and purification (sections 2.3.5.3 and 2.3.5.4).

Enzyme	Protein yield	Enzyme	Protein yield
LigE	114 mg·L^{-1}	LigF-NA	227 mg·L^{-1}
LigE179	253 mg·L^{-1}	LigF008	130 mg·L^{-1}
LigE283	101 mg·L^{-1}	LigF215	190 mg·L^{-1}
LigE491	80 mg·L^{-1}	LigF729	72 mg·L^{-1}
LigE760	314 mg·L^{-1}	LigF755	158 mg·L^{-1}
LigE889	184 mg·L^{-1}	LigF921	228 mg·L^{-1}
LigE915	497 mg·L^{-1}	LigF935	253 mg·L^{-1}
		LigF965	31 mg·L^{-1}

3.3.3 Characterization of putative β-etherases

Activities of the purified β-etherases were investigated with the model compound 2,6-MP-VG and determined by HPLC (section 2.3.6.1, Figure 3.7). All 13 selected β-etherases displayed the expected activity and cleaved the β-4-O-acryl ether bond of the lignin model compound (Table 3.5). LigE283, LigE889, and LigF008 even showed two to three times higher activities than their respective reference enzymes LigE and LigF-NA, while most of the remaining enzymes (LigE179, LigE491, LigE915, LigF729, LigF755, LigF921, LigF935, and LigF965) were similarly active. In contrast, LigE760 and LigF215 displayed only 37% and 6% activity, respectively, compared to LigE and LigF-NA.

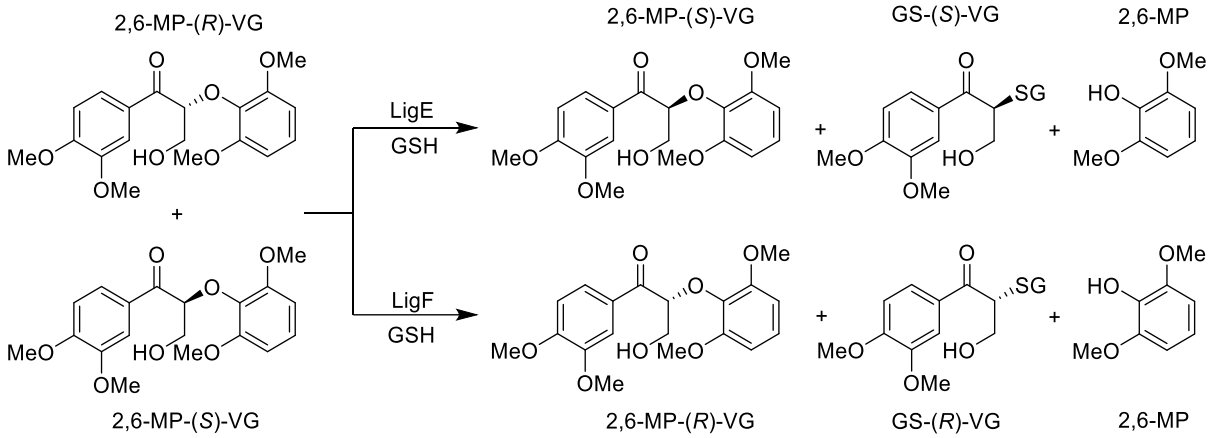


Figure 3.7 Reaction scheme for activity or selectivity assays of β-etherases with racemic 2,6-MP-VG as substrate (sections 2.3.6.1 and 2.3.6.2).

The enzymes' enantioselectivity in the kinetic resolution of 2,6-MP-VG was analyzed using chiral HPLC (section 2.3.6.2), following the decrease of 2,6-MP-(*R*)-VG and 2,6-MP-(*S*)-VG. All enzymes displayed absolute enantioselectivity with E-values higher than 200 (Table 3.5). In agreement with the selectivity of the reference enzymes, all LigE-type β-etherases were (*R*)-selective, whereas LigF-type enzymes converted the (*S*)-enantiomers of the lignin model substrate.

Table 3.5 Specific activity and enantioselectivity of the investigated enzymes in the cleavage of 2,6-MP-VG determined by HPLC (sections 2.3.6.1 and 2.3.6.2).

Enzymes	Specific activity [U·mg^{-1}]	Selectivity	E-value
LigE	0.65	(*R*)-selective	>200
LigE179	0.94	(*R*)-selective	>200
LigE283	1.86	(*R*)-selective	>200
LigE491	0.98	(*R*)-selective	>200

Enzymes	Specific activity [U·mg^{-1}]	Selectivity	E-value
LigE760	0.24	(*R*)-selective	>200
LigE889	1.67	(*R*)-selective	>200
LigE915	0.66	(*R*)-selective	>200
LigF-NA	2.06	(*S*)-selective	>200
LigF008	4.69	(*S*)-selective	>200
LigF215	0.12	(*S*)-selective	>200
LigF729	1.77	(*S*)-selective	>200
LigF755	2.20	(*S*)-selective	>200
LigF921	3.48	(*S*)-selective	>200
LigF935	2.92	(*S*)-selective	>200
LigF965	2.07	(*S*)-selective	>200

All enzymes were also characterized in respect to their temperature stabilities as well as pH and temperature optima (sections 2.3.6.3 and 2.3.6.4). In activity assays at different pH values using the fluorogenic model substrate MU-VG, all enzymes displayed activity in the range between pH 6 and pH 10 with an optimum clearly in the alkaline region at pH 9 (Figure 3.8 **A** + **B**). To exclude buffer effects during pH optima determination, the assay was also performed with a mixture of Tris and glycine buffer at pH 8.5, 9.0 and 9.5 giving the same result as with the single buffers.

Optimal reaction temperatures for activity were obtained in a range between 20 and 40 °C. Nevertheless, many enzymes still show relevant residual activity above 40 °C. For example, LigF935 has still 60% relative activity at 50 °C (Figure 3.8 **C** + **D**).

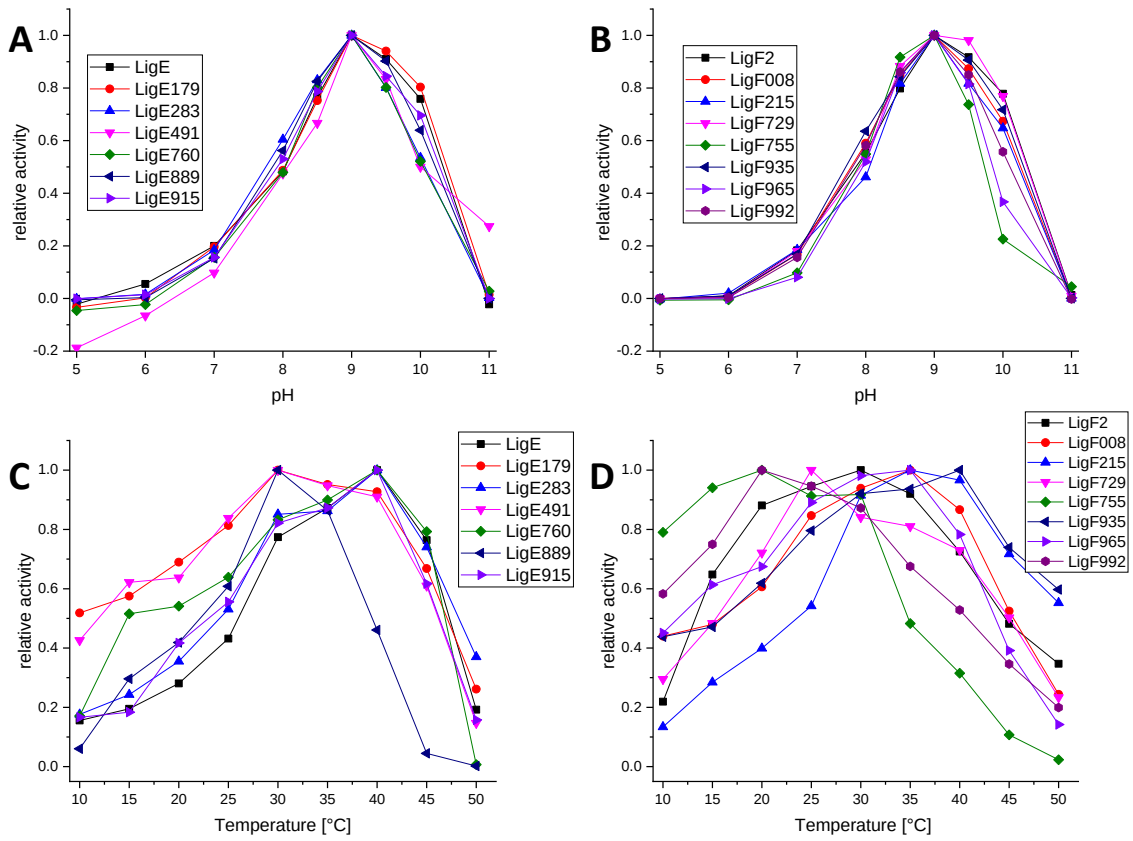


Figure 3.8 pH and temperature profiles of investigated β-etherases. The reactions for pH optima (**A** + **B**) were performed in 200 µL scale with fluorogenic substrate MU-VG, measuring the fluorescence of the released 4-methylumbelliferone at 450 nm (section 2.3.6.3). For determination of temperature profiles (**C** + **D**) the substrate VN-VG was used in an absorbance assay of 400 µL reaction volume, measuring the absorbance of the released vanillin at 360 nm (section 2.3.6.3). The highest activity for each enzyme was set to 1 to calculate relative activities.

The temperature stabilities of the enzymes were investigated using the thermofluor assay to determine individual melting temperatures (T_m). As a result, obtained T_m values ranged between 43 °C and 69 °C (Table 3.6). The T_m of LigF215 with 43 °C is significantly lower than the T_m values of most other tested β-etherases, which possess melting points above 50 °C. This lower thermal stability of LigF215 is in direct agreement with the observed low stability and precipitation of the enzyme encountered during purification.

Table 3.6: Temperature and pH optima as well as apparent melting temperatures of investigated β-etherases. pH optima were determined with the fluorogenic substrate MU-VG measuring the fluorescence of the released 4-methylumbelliferone (section 2.3.6.3), whereas temperature optima were determined using the substrate VN-VG in an absorbance assay (section 2.3.6.3). Apparent melting temperatures were determined using the thermofluor assay (section 2.3.6.4).

Enzyme	pH optimum	Temperature optimum [°C]	Apparent melting temperature [°C]
LigE	9.0	40	52.5
LigE179	9.0	30	53.8
LigE283	9.0	40	55.8
LigE491	9.0	30	51.3
LigE760	9.0	40	51.3
LigE889	9.0	30	50.3
LigE915	9.0	40	69.0
LigF-NA	9.0	30	62.8
LigF008	9.0	35	57.5
LigF215	9.0	35	42.8
LigF729	9.0	25	55.0
LigF755	9.0	20	48.5
LigF921	9.0	20	65.0
LigF935	9.0	40	67.2
LigF965	9.0	35	51.5

3.3.4 Sequence and structure-function analysis

The analysis of multiple sequence alignments of an enzyme family can reveal important information such as conserved amino acids or motifs, which often are essential for function or structure of the enzymes. Especially the sequence-structure-function relationship of β-etherases is not well understood yet. Here, the conserved small peptides, identified by the PPR algorithm, are a good starting point for sequence analysis.

As already mentioned, β-etherases consist of two domains, the N-terminal thioredoxin and the C-terminal helical domain. In the case of LigE-type enzymes, the majority of the conserved peptides identified by the PPR algorithm is found in the N-terminal thioredoxin domain (residues 1-82 in LigE), which also harbours the GSH-binding site. This high level of conserved amino acids in the thioredoxin domain is also observed in the corresponding webPRANK alignment. The conserved hexamer peptides of the LigE-group can be assembled into larger motifs with putative importance and function in LigE-type enzymes. In the thioredoxin domain, three motifs were identified: GxTx**S**PxVWxxxxAxxHKG, Rx**P**xIxDxG, and L**DS**WxIxExLD. The amino acids S21, P60, D71, and S72 (LigE numbering, highlighted in bold in previous

motifs) are strictly conserved and in the LigE crystal structure (PDB: 4YAN) seem to interact with co-substrate GSH (Figure 3.9, **A**).

In contrast, no long-conserved motifs could be identified in the C-terminal helical domain (residues 93 to 255 in LigE), but still several amino acids such as W105, W107, Y122, F142, and W197 are highly conserved. Unfortunately, the enzymes' exact substrate binding mode, except for GSH, is not known as no substrate-bound crystal structures of LigE and LigF could be obtained.[69] On the other hand, it is highly remarkable, that many solvent exposed hydrophobic and especially aromatic amino acids are highly conserved in the C-terminal domain of LigE-type enzymes. With lignin being an aromatic and highly hydrophobic substrate, such conserved hydrophobic and aromatic amino acids will likely be important for substrate binding.

To better understand which amino acids are involved in these interactions, substrate 2,6-MP-VG was docked into the crystal structure of LigE. Even though none of the resulting possible binding modes seemed to resemble the productive substrate conformation, as the distance between the GSH thiolate and the substrate's β-carbon and geometry was in each case not in agreement with an S_N-2 reaction mechanism, the conformation exhibiting the shortest S^--C_β distance (4.9 Å) was used for further analysis. In the docked structure, amino acids Y23 (thioredoxin domain), W107, Y122, F142 and W197 show direct interactions with the substrate (Figure 3.9, **B**). W107 and Y122 are strictly conserved, whereas only hydrophobic residues are found at the positions 23 (F,Y), 142 (F,L,W) and 197 (F,W). In addition to these residues with clear substrate interaction, conserved residue W105 occurs in every LigE-type enzyme but only flanks the substrate binding region while being part of the dimer interface in LigE.

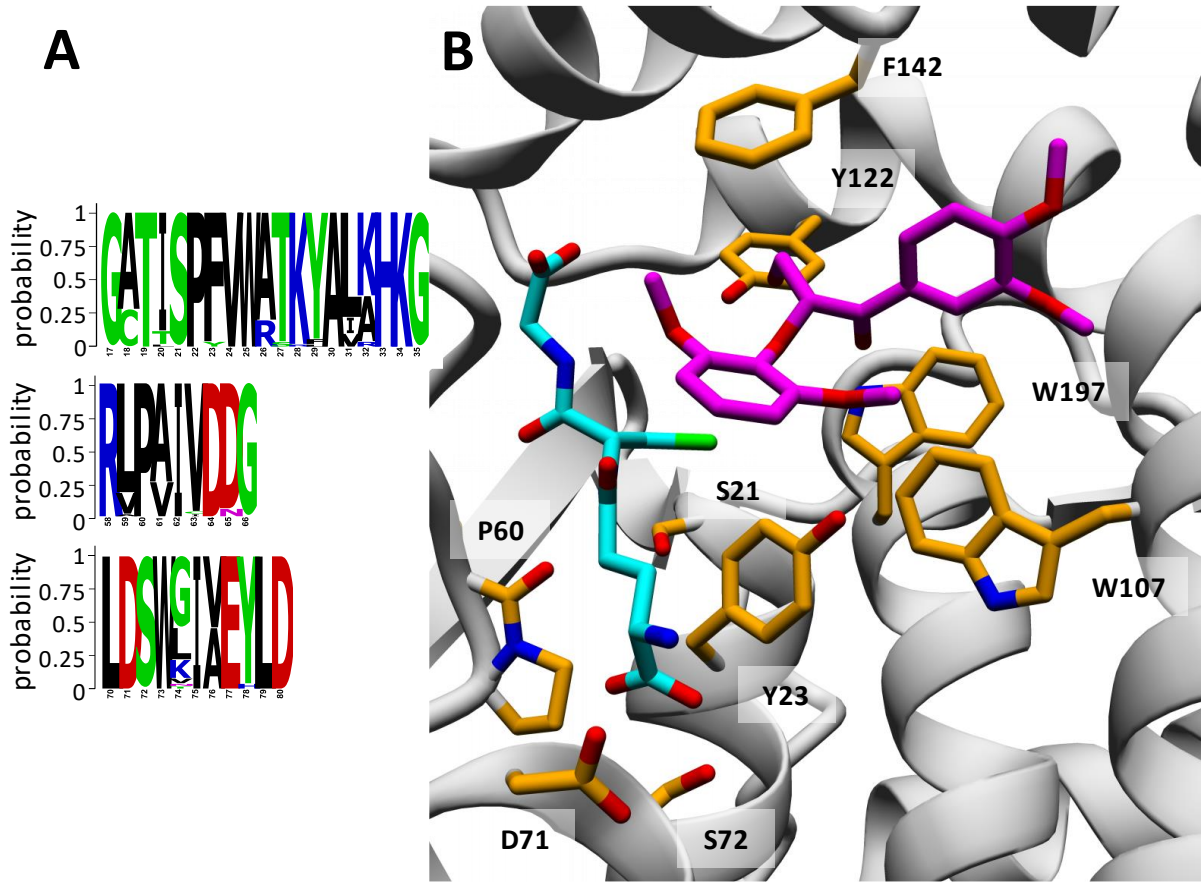


Figure 3.9 Visualization of conserved amino acids and sequence motifs in LigE enzymes. **A** sequence logos, visualized using WebLogo (section 2.3.3.3),[98] of the conserved amino acid motifs present in LigE-type enzymes. **B** active site of LigE (PDB: 4YAN) with the cofactor GSH and the docked substrate 2,6-MP-(*R*)-VG (section 2.3.3.4). The conserved amino acids S21, Y23, P60, D71, S72, W107, Y122, F142 and W197 surrounding the GSH or the substrate are shown in orange. GSH is colored in turquois, and 2,6-MP-(*R*)-VG in magenta.

In LigF-type enzymes, GSH-binding seems to be highly conserved as well. The crystal structure of LigF (PDB: 4XT0) indicates that amino acids N13, S14, and K16 (LigF-numbering, highlighted in bold in the following motif) of the thioredoxin domain (residues 1-76 in LigF) motif LYSFGPxA**NS**xK**P**, which is found in nearly all LigF-type enzymes, interact with the co-crystallized cofactor GSH (Figure 3.10 **A**). Only the LigF-type β-etherase from *Sphingomonadales* bacterium 39-62-4 (GenBank accession number OZA57515) does not share this motif. Also, S67 interacts with GSH and is part of motif TE**S**TVICEYLEDxxP, which is present in all LigF-type enzymes.

In contrast to the LigE-type etherases, long stretches in the helical domain (residues 93-242 in LigF) are highly conserved in LigF-type enzymes. In particular, the long motif **A**xM**R**xWT**K**WVDEYFCWCVSTxGW is striking merely due to its size. A part of this motif (A94, R97, K101) is located in the dimer interface and is very likely involved in the interaction of both monomers (Figure 3.10, **A** + **C**).

As for LigE-type enzymes, the exact substrate-binding mode of LigF-type enzymes is still unknown. Docking substrate 2,6-MP-VG into the crystal structure of LigF yielded a conformation that likely resembles the productive binding mode based on the S^--C_β distance and substrate geometry (i.e., the leaving group is on the opposite site of the β-carbon relative to the approaching thiolate). Inspection of this substrate-bound structure revealed the strictly conserved amino acids F8, W109, V111, S112, W116, and W149 as well as positions 120 (I,V) and 200 (I,V) as interaction partners of the substrate. Residues W109, V111, S112, and W116 belong to the second half of the aforementioned motif AxMRxWTKWVDEYFC**W**C**VS**TxG**W**, underlining its importance for substrate binding as well as dimerization of LigF-type enzymes (Figure 3.10, **B**). Additionally, several hydrophobic amino acids between positions 179 and 209, such as L179, L185 and L192, are conserved but do not form a closed motif.

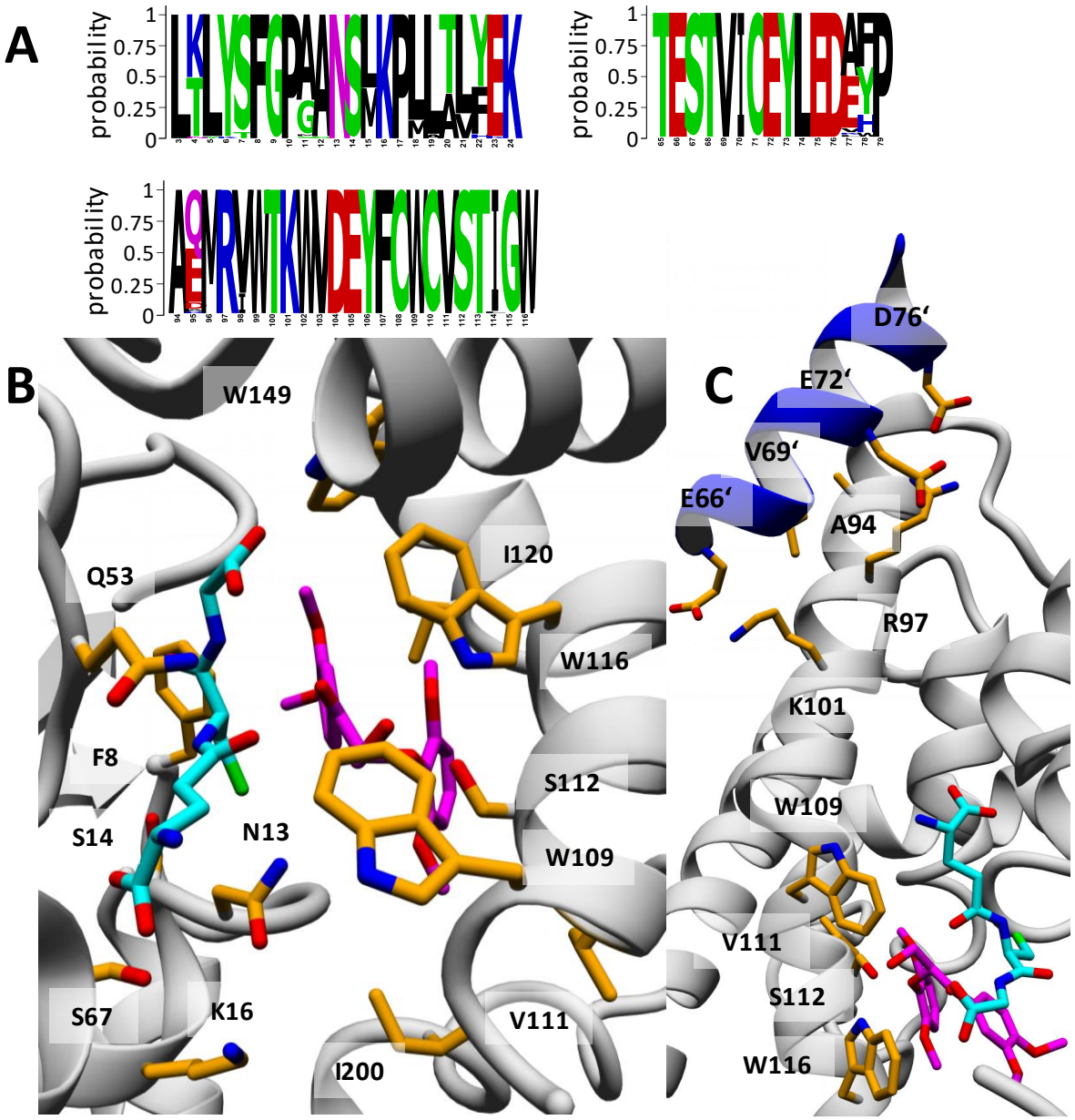


Figure 3.10 Visualization of conserved amino acids and sequence motifs in LigF-type enzymes. **A** Sequence logos, visualized using WebLogo (section 2.3.3.3),[98] of the conserved amino acid motifs present in LigF-type enzymes. **B** Active site of LigF (PDB: 4XT0) with the cofactor GSH and the docked substrate 2,6-MP-(*S*)-VG (section 2.3.3.4). **C** Dimer interface of LigF with chain A shown in grey and chain B in blue. Conserved amino acids interacting either with GSH (F8, N13, S14, K16, Q53, S67) or the substrate (W109, V111, S112, W116, I120, W149, I200) as well as conserved residues in the dimer interface (A94, R97, K101) are visualized in orange. GSH is coloured in turquois and 2,6-MP-(*S*)-VG in magenta.

This sequence and structure analysis with the increased set of β-etherases will be a good starting point for further functional analysis of these enzymes to better understand their catalytic mechanism and the functional role of conserved amino acids.

3.3.5 Phylogenetic analysis

To analyse the phylogenetic relationship of β-etherases and homologous glutathione S-transferases another BLASTP search was performed using all previously known and the 13 herein confirmed β-etherases as queries and limiting the search to 1,000 hits per query sequence (nr database of GenBank, release 226). The resulting 5,026 sequences (1,118 found by BLASTP based on LigE-type query sequences and 3,908 found by BLASTP based on LigF-type query sequences after removal of double entries) were used to construct a phylogenetic tree based on average linkage (UPGMA) using the MAFFT server (Figure 3.11).

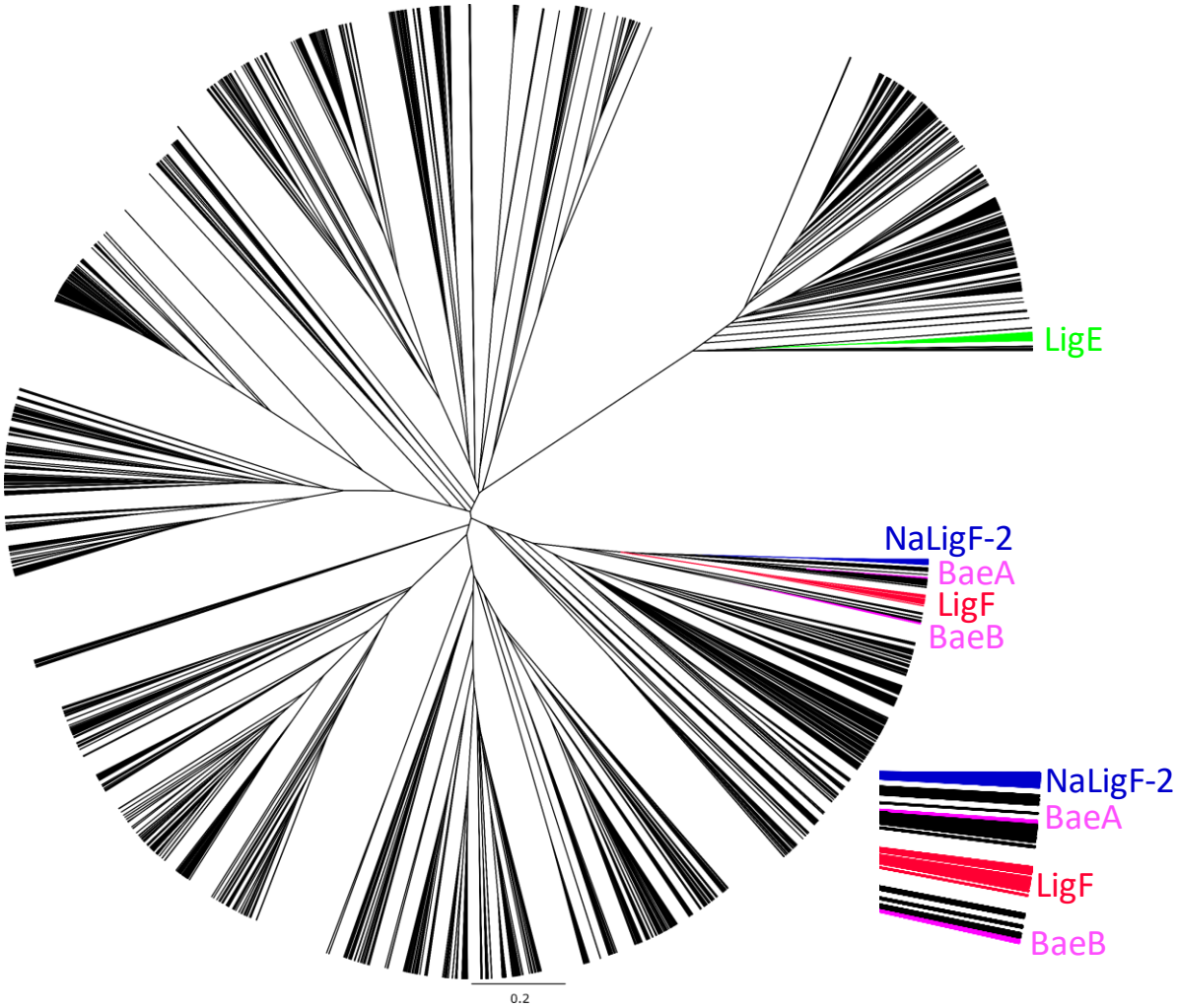


Figure 3.11 Phylogenetic analysis of β-etherases and homologous sequences. The phylogenetic tree was generated using MAFFT[97] (section 2.3.3.3) based on 5,026 protein sequences obtained after standard BLASTP searches with all previously known and 13 herein confirmed β-etherases as queries (limited to 1,000 sequence hits per query; all sequences were combined and double entries were removed, query genes are listed in section 6.4). LigE-type sequences are colored in green, whereas LigF-type sequences are colored in red. Sequences grouped together with NaLigF-2 by the PPR algorithm are colored in blue. The phylogenetic branches containing both sequences encoding the recently described hetero-dimeric β-etherase BaeAB (BaeA, BaeB) from *Novosphingobium aromaticivorans* are colored in magenta.

This phylogenetic analysis confirms that LigE- as well as LigF-type sequences are clearly separated from each other. Also, LigF-type enzymes and NaLigF-2 are placed on separate branches of the phylogenetic tree. The latter observation is in agreement with the significantly lower sequence homology of NaLigF-2 to other known LigF-type enzymes and the placement of NaLigF-2 in a different PPR group. Hence, LigF-type and NaLigF-2-type enzymes can be distinguished among bacterial (*S*)-selective β-etherases.

The phylogenetic branches covering LigE- and LigF-type enzymes contain the same sequences, which were also clustered together by the PPR algorithm. Moreover, the same is true for NaLigF-2, which was clustered together with 52 other homologs within the PPR analysis, performed for the identification of novel β-etherases (section 3.3.1). The same enzymes grouped by the PPR algorithm together with NaLigF-2 are found on one separated phylogenetic branch together with NaLigF-2 (GenBank accession numbers of the enzymes in the NaLigF-2 PPR group are listed in section 6.4). Hence, the PPR result is congruent with the phylogenetic placing of β-etherases. The same phylogenetic arm carrying LigF- and NaLigF-2-type enzymes contains also the sequences coding for the recently described (*R*)-selective hetero-dimeric β-etherases BaeAB. Hence, all three β-etherase types seem to originate from a common ancestor.

The phylogenetic analysis was also performed on a bigger dataset generated by BLASTP searches with 5,000 hits per query, resulting in 30,000 sequences in total. This analysis revealed the same phylogenetic relationships as the one displayed in Figure 3.11 and is therefore not shown.

The phylogenetic data are also a great starting point to investigate putative enzymes with potential activity for β-*O*-4 aryl ether cleavage beside the classic LigE- and LigF-type enzymes. The NaLigF-2 homologs also grouped together by the PPR algorithm, as well as enzymes from LigE- and LigF-neighbouring branches of the phylogenetic tree could reveal interesting activities and enzymatic properties. Especially the subtree with LigF-, NaLigF-2- and BaeAB-type enzymes seems to be very interesting, as (*R*)- and (*S*)-selective β-etherases are located here.

3.4 Protein engineering of LigG-TD

In the pathway of GSH-dependent lignin degradation, the GSH-adduct formed by β-etherase catalysis is further converted by glutathione lyases. At the starting point of this project only

the three Omega-class glutathione lyases LigG, LigG-NS, and LigG-TD were described. While for an efficient lignin degradation the cleavage of both enantiomers of the GSH-adduct is required, Omega-class glutathione lyases only convert the (*R*)-enantiomer of the GSH-adduct with high activity. Enzymes able to convert the (*S*)-enantiomer effectively were missing. Furthermore, for a preparative process it would be interesting to convert both enantiomers of the GSH-adduct using one unselective enzyme.

LigG-TD was the least selective published glutathione lyase at this time and was chosen as starting point for protein engineering using a semi-rational mutagenesis approach with the goal to generate a highly active as well as unselective glutathione lyase. To accomplish this, the crystal structure of LigG-TD and a high-throughput assay for activity screening were required.

3.4.1 Crystallization of LigG-TD

LigG-TD was crystallized in collaboration with Prof. Wulf Blankenfeld and Dr. Christina Diederich from the Helmholtz Center for Infection Research in Braunschweig. Christina Diederich solved the structure as part of her PhD thesis.

Protein crystals of LigG-TD were successfully obtained with and without co-crystallization of GSH (Figure 3.12).

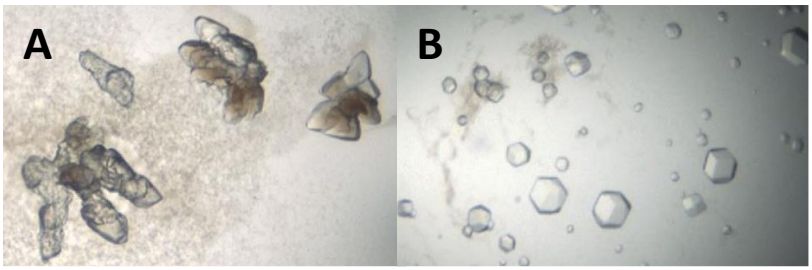


Figure 3.12 Protein crystals of LigG-TD. **A** 20 mg·mL^{-1} LigG-TD using the conditions JCSG Core I A1 (20wt% PEG 8000, 0.1 M CHES pH 9.0) after 13 days. **B** 10 mg·mL^{-1} LigG-TD with 10 mM GSH using the conditions JCSG Core I B3 (10 wt% PEG 8000, 0.1 M HEPES pH 7.5) after 2 days.

The best crystals diffracted up to 2.13 Å resolution for apo-LigG-TD and 2.8 Å resolution for LigG-TD containing GSH. The overall structure of LigG-TD is similar to the structure of LigG (Figure 3.13, **A**), with a RMSD (root-mean-square deviation, the average difference between two protein structures) of 1.153 Å (aligned and calculated with MUSTANG, implemented in Yasara) for LigG and LigG-TD co-crystallized with GSH. This confirms that also LigG-TD belongs to the Omega-class of GSTs.

Comparison of the apo-structure of LigG-TD to its structure containing GSH reveals no significant conformational changes, induced by GSH-binding (Figure 3.13, **B**). The RMSD of both structures is 0.654 Å.

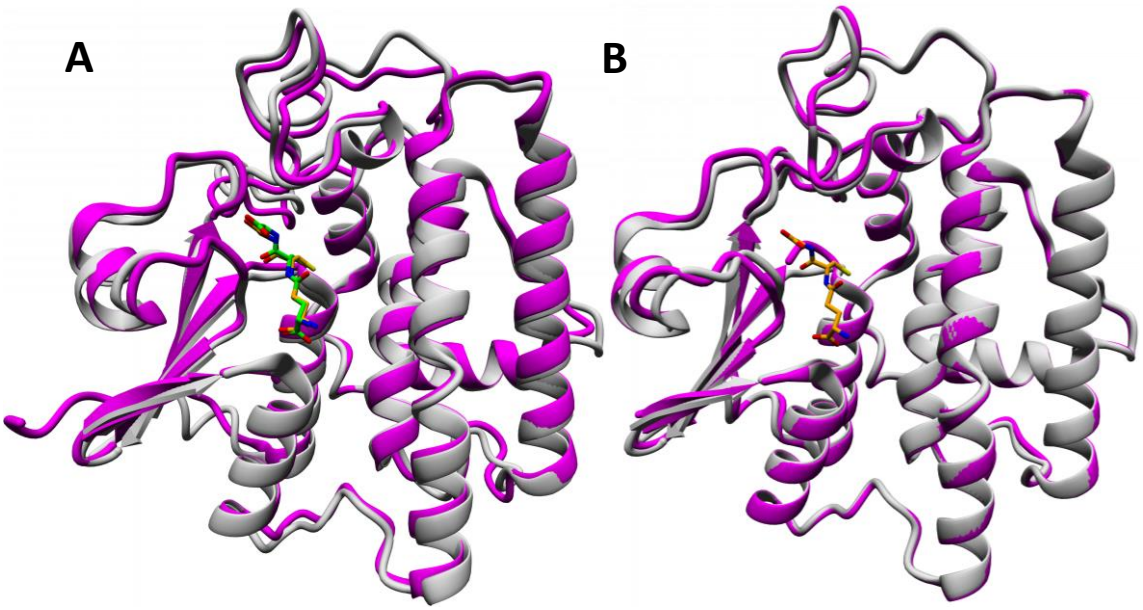


Figure 3.13 Superposition of LigG (PDB: 4g10, LigG magenta; GSH green) and LigG-TD (LigG-TD gray; GSH orange) co-crystallized with GSH (**A**) as well as apo-LigG-TD and LigG-TD co-crystallized with GSH (**B**, apo-LigG-TD magenta LigG-TD gray, GSH orange).

Beside the similarity of the overall structure, the amino acid composition in the active sites of LigG and LigG-TD differs (Figure 3.14), resulting in different activities and stereoselectivities of both enzymes. How this difference in amino acid composition is causing the higher activity of LigG and the lower stereoselectivity of LigG-TD is not known.

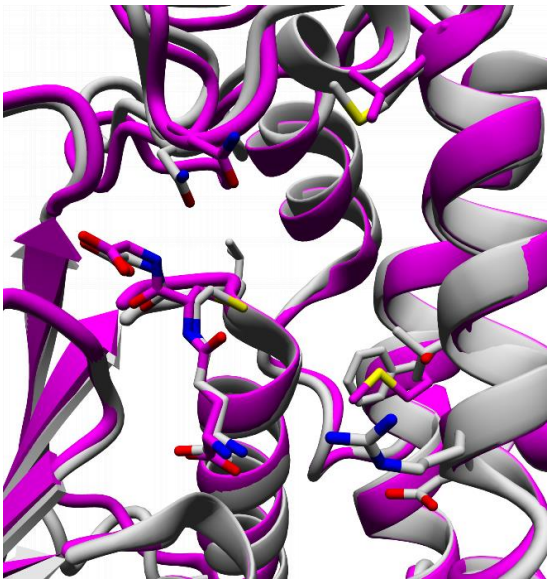


Figure 3.14 Superposition of the active site of LigG (PDB: 4g10, magenta) and LigG-TD (gray) co-crystallized with GSH.

3.4.2 Substrate docking

Based on the crystal structure of LigG-TD, the number of amino acids for the mutagenesis should be limited. Therefore, both enantiomers of GS-VG were first docked in the crystal structure of LigG-TD using AutoDock Vina implemented in Yasara. As the exact reaction mechanism of glutathione lyases is not known yet, there are still two possibilities for the localization of the GSH-adduct in the active site. LigG-TD either binds only the GSH-adduct or GSH and the GSH-adduct at the same time.

Pereira et al. suggested a two-step mechanism for the Omega-class glutathione lyase LigG (in the following LigG-numbering is used; C15 and Y113 in LigG are the corresponding residues of C12 and Y112 in LigG-TD).[71] In the first step, the substrate GS-VG is bound in the enzyme active site mediated by Y113. Then, C15 performs a nucleophile attack on the sulphur of the thioether bond resulting in GS-VG cleavage. VG is released and a covalent enzyme-GSH intermediate is formed. In the next step, this intermediate is cleaved by nucleophilic attack of a second GSH molecule with GSSG as product.

Another possible mechanism was suggested by Kontur et al. for glutathione lyases belonging to the Nu-class of GSTs.[60] In this mechanism, the GSH-adduct and GSH are bound at the same time in the enzyme active site. The sulphur of the free GSH performs a direct nucleophilic attack on the sulphur of the GSH-adduct resulting in the cleavage of the GSH-adduct and release of GSSG. Since both enzyme classes perform the same reaction, it is likely that the main characteristics of the catalytic mechanism are also the same. But the active sites of Omega-class and Nu-class glutathione lyases differ highly (see section 1.8.4). Therefore, the one-step mechanism postulated by Kontur et al.[60], with simultaneous binding of the GSH-adduct and GSH and direct nucleophilic attack of GSH on the GSH-adduct, was analysed as a possible mechanism for LigG catalysis as well. In this new mechanism also the amino acids C15 and Y113 are involved, since their participation in the LigG catalysis is already proven.[71] How plausible both mechanism are, has to be elucidated in the following by substrate docking and experimentally. In Figure 3.15, the reaction schemes of both possible mechanisms, the two-step mechanism postulated by Pereira et al.[71], and the one-step mechanism according to Kontur et al.[60], are illustrated.

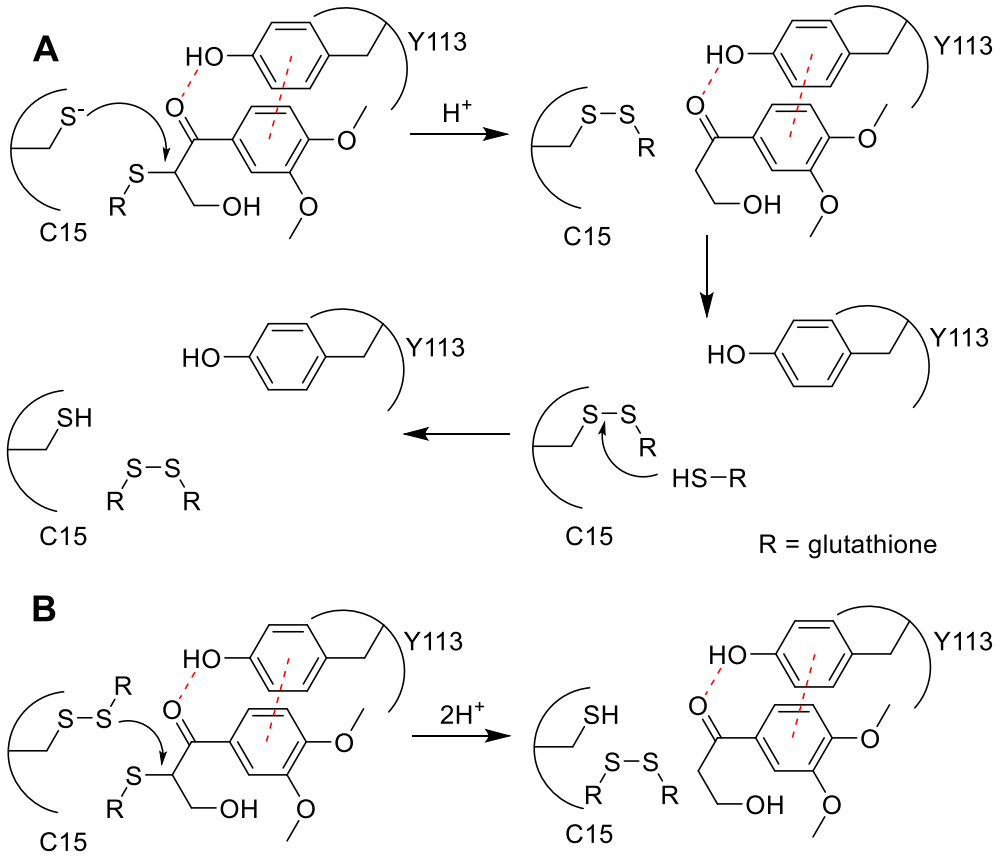
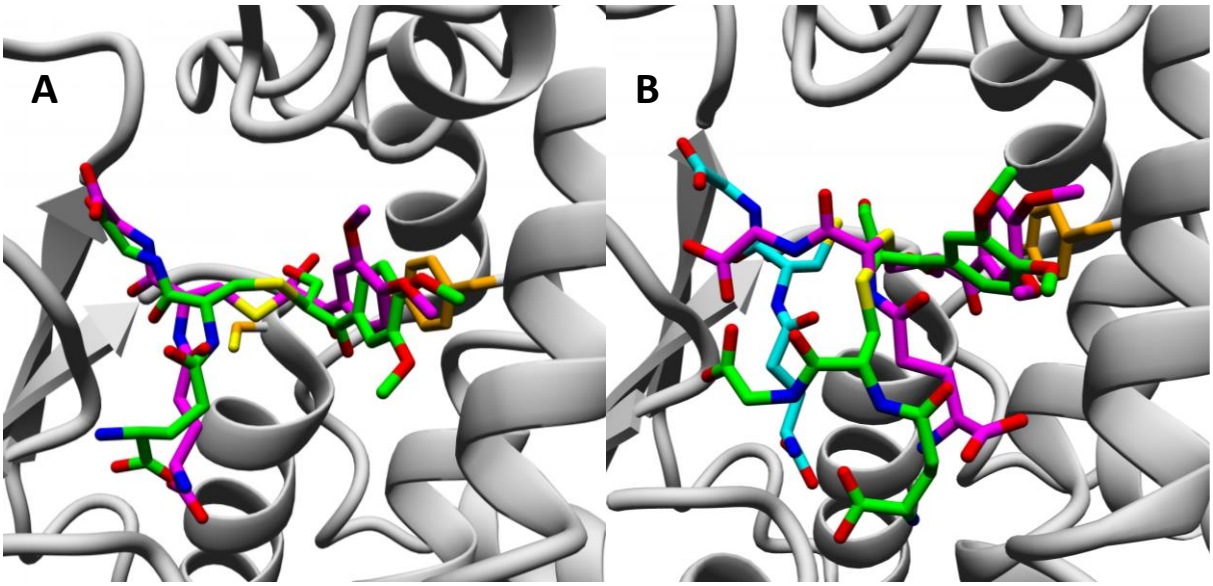


Figure 3.15 Reaction schemes of the two possible reaction mechanisms of glutathione lyases in the cleavage of GSH-adducts, postulated by Pereira et al. (**A**) and according to the mechanism postulated by Kontur et al. (**B**).[60,71] The numbering of the amino acids is according to LigG amino acid numbering.

Both enantiomers of GS-VG were docked into the crystal structure of LigG-TD with and without bound GSH. The docking results are shown in Figure 3.16. In the docking attempts, substrate binding modes were obtained, that would agree with both possible catalytic mechanisms. Furthermore, the docking studies also explain the *R*-selectivity of LigG-TD. With a sulphur-sulphur distance (see mechanisms in Figure 3.15) in all dockings of the (*R*)-enantiomer (Figure 3.16 **A** + **B** magenta) below 4 Å (3.675 Å and 3.569 Å, respectively) and the right geometry of the VG-leaving group, the structures agree with typical criteria of an S_N2 mechanism. In contrast to that the sulphur-sulphur distance in the best docking poses of the (*S*)-enantiomers (Figure 3.16 **A** + **B** green) is above 4 Å (5.534 Å and 6.186 Å, respectively) as well as the geometry of VG-leaving group does not fit to the S_N2 criteria. This correlates with the lower conversion of (*S*)-configurated substrate by LigG-TD.

In the docking of GS-(*R*)-VG into the crystal structure of LigG-TD co-crystallized with GSH (Figure 3.16 **B** magenta) only in the first attempts with 200 docking runs a productive docking pose was obtained. In later attempts with 999 docking runs and equal settings this pose was

not found again. Nevertheless, the docking pose obtained in the docking 200 runs explains the LigG-TD catalysis well and is therefore used further.



A: GS-VG docked according to the two-step mechanism postulated by Pereira et al.[71]

B: GS-VG docked according to the one-step mechanism postulated by Kontur et al.[60]

Figure 3.16 Results of LigG-TD substrate docking. Using AutoDock Vina (implemented in Yasara), the substrates GS-(*R*)-VG and GS-(*S*)-VG were docked in the crystal structure of apo-LigG-TD (**A**) as well as the LigG-TD structure containing GSH (**B**). GSH is colored in turquoise, GS-(*R*)-VG in magenta, GS-(*S*)-VG in green. The catalytic active amino acids C12 and Y112 are colored in orange. The docking (section 2.3.3.4) was performed with 999 docking runs, except for the docking of GS-(*R*)-VG in LigG-TD co-crystallized with GSH (**B**, colored in magenta). In this case 200 docking runs were used.

3.4.3 Development of a high-throughput screening assay

A bottleneck in protein engineering is usually the screening of mutant libraries for variants with improved properties. The standard activity assay procedures for glutathione lyases are based on HPLC, which is too time-consuming for an efficient engineering process. Instead, in high-throughput screening (HTS) assays one reaction component is usually quantified directly or indirectly using spectrophotometric or fluorescent methods.

The substrates and products of the glutathione lyase reaction do not exhibit a specific absorbance or fluorescence, which could be used for quantification. Therefore, an indirect assay had to be used. To make the assay more general, the attempts focused on the GSH and GSSG. Furthermore, for the quantification of GSH and GSSG different methods have been published, which were tested.

3.4.3.1 Detection of GSSG using GSH-specific dyes

The photometric detection of GSH is usually based on the derivatization of the GSH thiol group with GSH-specific dyes. The photometric properties of the dye are altered by this reaction, which is used for the detection. Since GSSG does not have a free thiol group, it has to be reduced before. By this way either the GSSG or total GSH (GSH + GSSG) concentration can be measured. If explosively the GSSG concentration should be determined, the free GSH has to be removed by derivatization first.

GSSG detection assay published by Rahman et al.

One assay for the quantification of GSSG was published 2006 by Rahman et al.[103] The first step in this GSSG detection assay is the derivatization of reduced GSH by 2-vinylpyridine, for the selective determination of GSSG. Then, GSSG is reduced by the glutathione reductase and reduced GSH is formed again. This GSH is detected based on a reaction with Ellman's reagent [DTNB, 5,5'-dithiobis-(2-nitrobenzoic acid)]. In this reaction a GSH-adduct and 2-nitro-5-thiobenzoate (TNB), which absorbance (A_{412nm}) is measured, are formed. Since the glutathione reductase is also active on the GSH-adduct, it is not possible to determine the GSSG concentration based on the end concentration of TNB. Therefore, the kinetic of the TNB formation is used, which is dependent on the concentration of GSSG and glutathione reductase. By linear regression of the TNB formation rate and the comparison with reactions using GSSG standards as substrate the GSSG concentration can be determined. The principle of the assay is shown in Figure 3.17.

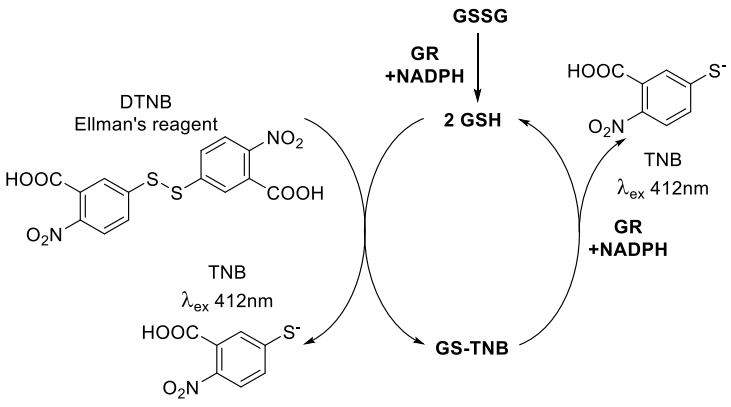


Figure 3.17 Reaction scheme of the GSH detection assay published by Rahman et al. Adapted from Rahman et al.[103]

The GSH detection assay could be established successfully. The GSSG concentration of different samples could be determined with the help of GSSG calibration curves (Figure 3.18, **A** and **B**). The determination of the GSSG concentration in a 1 mL glutathione lyase reaction led to comparable results to the determination of VG with HPLC (GSSG and VG are in formed equimolar amounts in the glutathione lyase reaction, Figure 3.18 **C**)

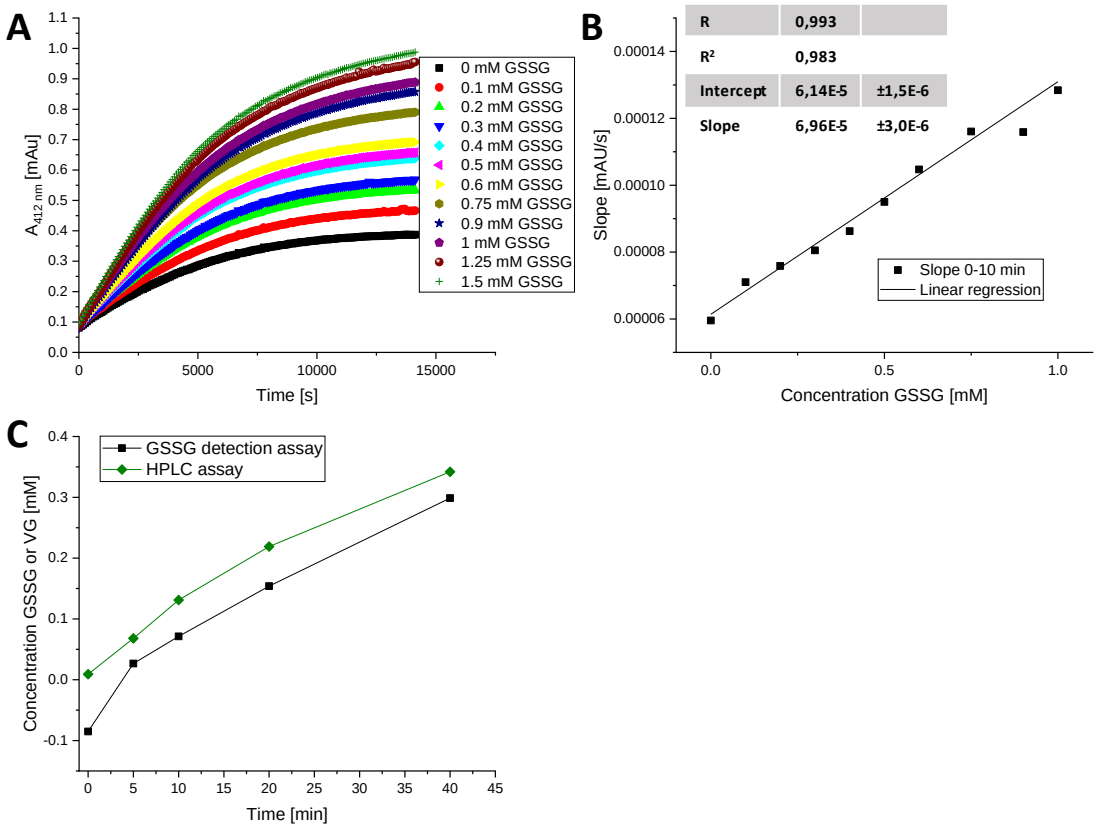


Figure 3.18 Establishing of the GSH detection assay by Rahman et al.[103] **A** Measurement of GSSG standards over the time, **B** calibration curve generated with these GSSG standards, **C** comparison of GSSG concentration measured with the GSH detection assay and the VG concentration measured with HPLC in a 1 mL glutathione lyase reaction over time.

One drawback of the assay is the time-consuming workflow. For example, the derivatization of free GSH requires 1 h of incubation. Furthermore, the assay determines the GSSG concentration offline and discontinuously. In consequence, per reaction 3 to 4 samples have to be analyzed for the determination of the enzyme activities.

Another drawback is the kinetic readout of the assay. The reaction rate of this assay depends on the GSSG and the glutathione reductase concentration. The glutathione reductase is a very active enzymes and has due to this also a big influence on the assay reaction rate. Since it is due to pipetting errors hardly possible to prepare glutathione reductase stock solutions repetitively with the exact same concentration, only samples prepared with the same

glutathione reductase stock solution are comparable and also the calibration curve has to be prepared for every glutathione reductase stock solution individually. Due to these reasons the GSH detection assay is not suitable for an HTS of mutant libraries. In conclusion, the assay works well for a limited number of samples, however, due to the time-consuming and work-intensive workflow of the assay, a high sample throughput is hardly possible.

GSH detection using other dyes

In addition to the assay reported by Rahman et al., two further dyes for GSH detection were tested according to reports by Das et al. and Yi et al.[94,104] The respective reactions of the dyes with GSH are shown in Figure 3.19. In case of L_2, the corresponding GSH-adduct is reported to exhibit an absorbance maximum around 450 nm, whereas the respective GSH-adduct with the coumarin derivative exhibits an emission maximum at 465 nm.[94,104]

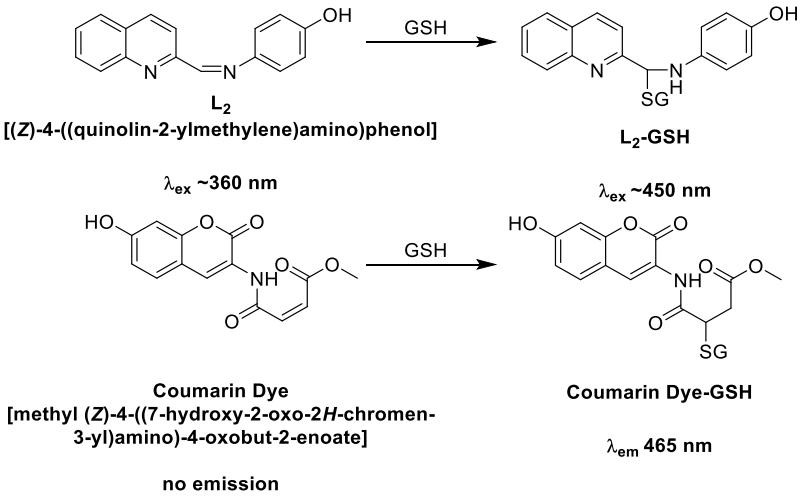


Figure 3.19 Molecular reactions of L_2 and coumarin dye with GSH. Shown are the chemical structures and reported change in absorbance or emission maxima.[94,104]

Although the chemical synthesis of L_2 was successful, as confirmed by NMR and mass spectrometry, its incubation with and without GSH for 3 h did not give the expected result. Instead of a shift in the absorbance maximum, a general reduction of the L_2 absorbance was observed. A possible reason for this reduction in absorbance is the degradation of the L_2 dye. This might be caused by hydrolysis of the shift base under the reaction conditions or by photo degradation. In Figure 3.20 the absorbance spectra of L_2 incubated with and without GSH at time point zero and after 3 h is shown. The absorbance spectrum of L_2 at time point zero is in agreement with respective spectra published by Das et al.[94]

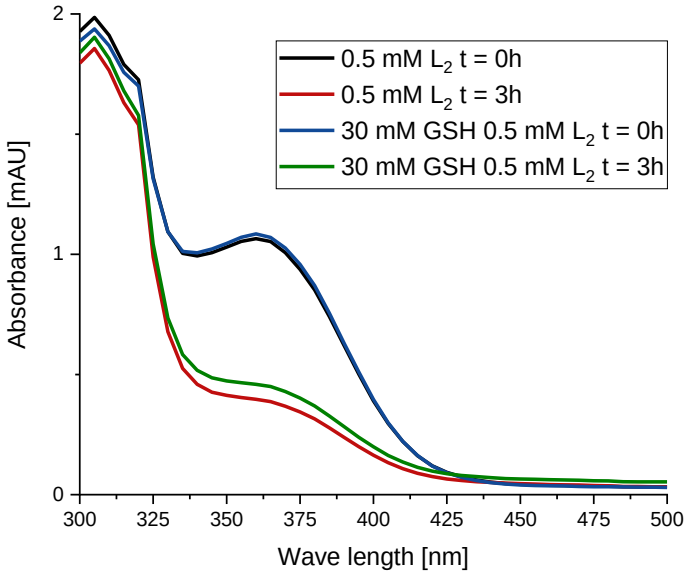


Figure 3.20 Absorbance spectra of L_2 with and without GSH incubation. 0.5 mM L_2 were dissolved in a DSMO-Hepes buffer (50 mM, pH 7.4) mixture (7:3 = DMSO : Hepes buffer) with and without 30 mM GSH. The absorbance spectra were measured directly after mixing L_2 and GSH (t = 0h) as well as after 3h (t = 3h) incubation.

Also, the synthesis of the coumarin dye was tested following the method of Yi et al. (Figure 3.21). After hydrolytic cleavage of the acetyl protection groups, however, no deprotected coumarin derivate was observed. Therefore, the synthesis was not continued and the coumarin dye could not be tested for GSH detection.

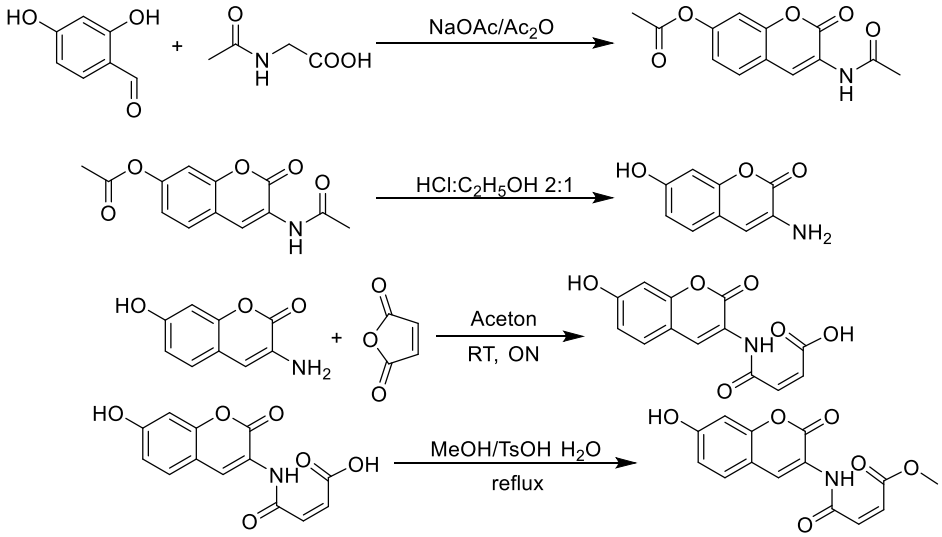


Figure 3.21 Synthesis scheme for the preparation of the coumarin dye according to Yi et al. Adapted from Yi et al.[104]

3.4.3.2 Glutathione reductase based high-throughput assay

Since the direct detection of GSH with fluorogenic or chromogenic dyes was not successful, an alternative assay was required to determine the activity of glutathione lyases. Combination of the glutathione lyase reaction with reduction of the resulting GSSG by the NADPH-dependent glutathione reductase offers the possibility to detect the GSSG formation by the NADPH oxidation. Per molecule of GSSG, one molecule of NADPH is consumed. Therefore, the NADPH consumption by the glutathione reductase is equimolar to the production of VG and GSSG in the glutathione lyase reaction (Figure 3.22).

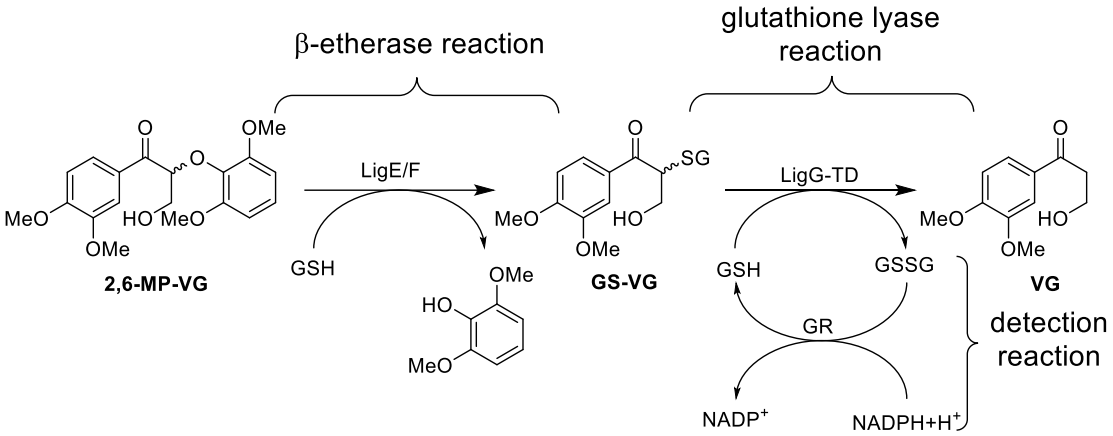


Figure 3.22 Reaction scheme of the glutathione reductase based HTS assay for the determination of glutathione lyase activity.

First step in the development of the glutathione reductase based HTS assay for the determination of glutathione lyase activity was to test different strategies for the cell lysis. In these experiments significant differences in the cell lysis of *E. coli* BL21 (DE3) Gold containing LigG-TD using lysozyme or B-PER were observed (Figure 3.23). Obviously, the addition of lysozyme did not efficiently lyse the cells under the used conditions and no production of GSSG was observed. The reaction curve is comparable to the control reaction using *E. coli* BL21 (DE3) Gold cells containing empty vector, whereas in the case of *E. coli* BL21 (DE3) Gold containing LigG-TD cells lysed with B-PER a clear decrease of NADPH in the glutathione reductase reaction was detected. Therefore, the cell lysis with B-PER was used for all further assay experiments.

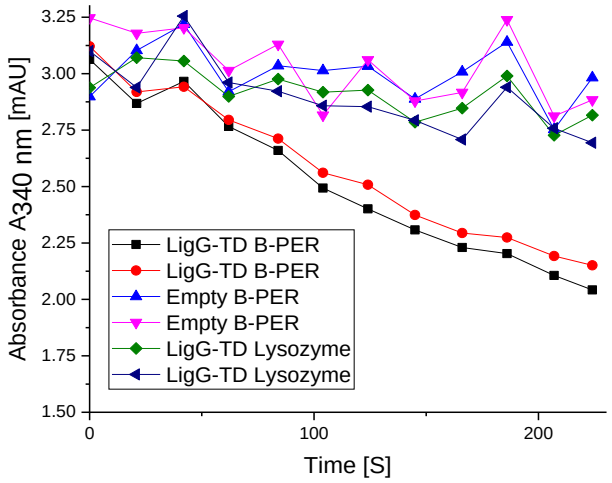


Figure 3.23 Test experiments for the development of a glutathione reductase based HTS assay to detect GSSG formed in the glutathione lyase reaction. The *E. coli* BL21 (DE3) cells with pET28a vectors were grown as descried in section 2.3.5.3. The cell pellets were lysed using either 200 µL B-PER for 20 min or 1 mg·mL^{-1} lysozyme for 1 h. The 200 µL reactions contained 0.8 mM GS-VG, 2 mM GSH, 100 mM Tris/HCl pH 8.5, 1.2 mM NADPH, 1 µL glutathione reductase and 25 µL cell lysate. The absorbance was measured at 340 nm. Visible is a consumption of NADPH in the samples containing LigG-TD, with cells lysed by B-PER, whereas the NADPH absorbance hardly decreases in the empty vector controls as well as in samples containing LigG-TD with cells lysed using lysozyme.

Additionally, the reaction pH had to be fine-tuned. Glutathione lyases have their pH optimum at pH 9, whereas the supplier of the used glutathione reductase recommends a reaction pH of 7.[59] The best compromise for the combination of both enzymes turned out to be pH 8.5. This pH is still near the pH optimum of LigG-TD, while the glutathione reductase is still highly active as well. After 40 min of incubation at pH 8.5 and room temperature, which is a sufficient time scale for the preparation and execution of the activity assay, no loss of glutathione reductase activity was observed. In contrast, at pH 9 the activity of the glutathione reductase is reduced (Figure 3.24).

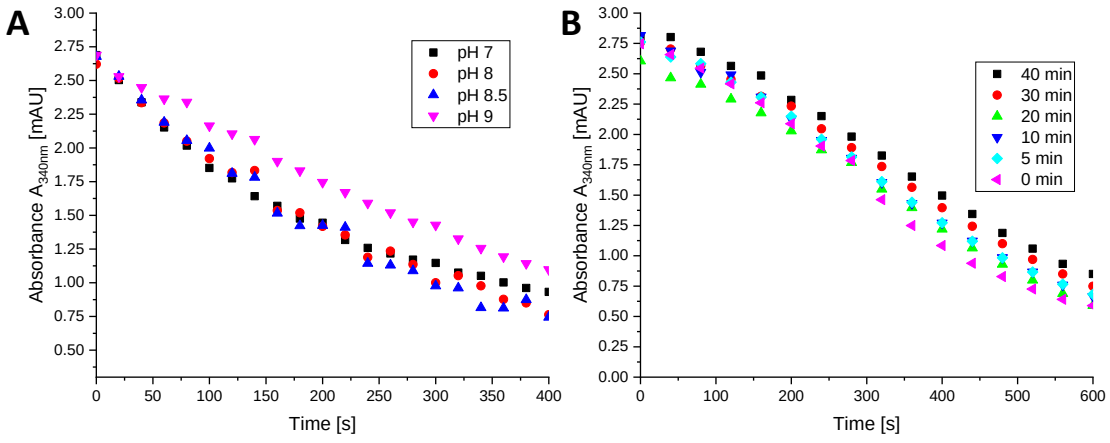
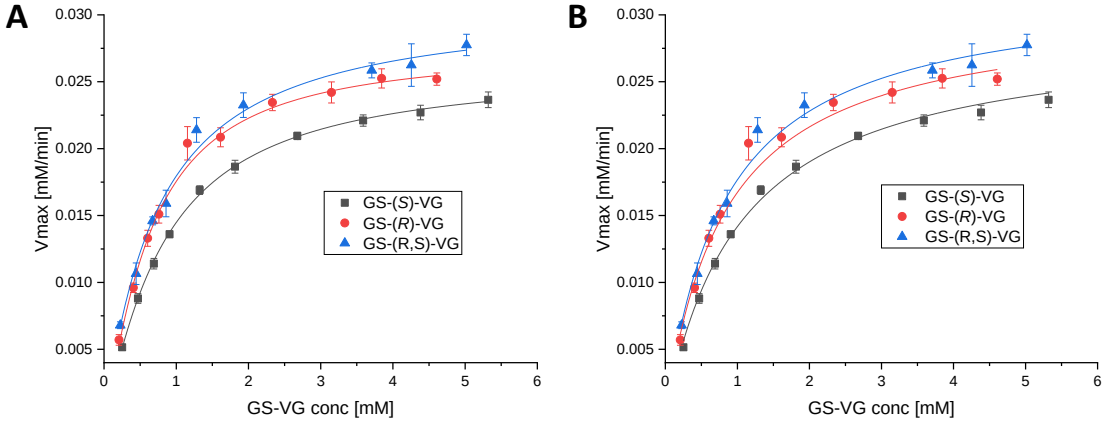


Figure 3.24 Optimization of glutathione reductase reaction. **A** Glutathione reductase reaction at different pH. **B** Glutathione reductase reaction after incubation of glutathione reductase at room temperature for a certain time.

The reactions were performed analog to the reactions described in section 2.3.7.4. The reaction scale was 200 µL with 100 mM buffer, 1.2 mM NADPH, 1 µL glutathione reductase and 0.8 mM GSSG as substrate.

Next important factor was the substrate concentration, which has a big impact on the enzyme activity if it is below the K_M value (Michaelis constant) or K value (Hill equation), respectively. This K-values represents the substrate concentration where the enzyme reaction rate is half V_{max} (maximum enzyme reaction rate). Therefore, the kinetic parameters for LigG-TD were determined, to evaluate which substrate concentration should be used in the glutathione reductase based HTS assay for LigG-TD. The K and K_M value for LigG-TD in the cleavage of GS-VG are between 0.6 and 1 mM, depending on the fitting model and the enantiomer of the substrate (Figure 3.25). The Hill equation fits the data points better than the Michaelis-Menten fit, indicated by the R^2 values. Due to the homo-dimeric form of LigG-TD interactions between the two monomers can lead to a cooperative binding and catalysis behavior, which would explain that the Hill equation fits the data points better than the Michaelis-Menten fit. Nevertheless, the cooperative effect in the case of LigG-TD is rather low, since the highest n-value, which is an indicator for the cooperativity of the enzymes, was observed in the cleavage of GS-(*R*)-VG by LigG-TD with a value of 1.265. The maximal n-value for a dimer, which could be reached, is 2, whereas 1 means no cooperative effect.

The substrate concentration in activity assays should be in a region where small changes in the substrate concentration do not have a big impact to the enzyme activity. In case of LigG-TD 4 mM for racemic GV-VG as well as 2 mM for GS-(*R*)-VG and GS-(*S*)-VG were chosen. At this substrate concentration small differences in the substrate concentration will not cause high differences in reaction rate, since they are above or equal to two-times the K_M or K value, respectively, and the enzyme activity is nearly saturated (Figure 3.25). Higher substrate concentrations would be beneficial, however, are due to the solubility of 2,6-MP-VG in the etherase reaction (Figure 3.22) hardly possible.



A) Hill $y=V_{max}\cdot X^n/(K^n+X^n)$

	GS-(*S*)-VG	GS-(*R*)-VG	GS-(*R*,*S*)-VG
V_{max} [mM/min]	0,026 ± 0.0004	0.028 ± 0.0007	0.031 ± 0.001
K_{cat} [1/s]	56.95 ± 0.88	61.33 ± 1.53	67.90 ± 0.22
K [mM]	0.833 ± 0.027	0.641 ± 0.037	0.732 ± 0.063
n	1.185 ± 0.031	1.265 ± 0.074	1.084 ± 0.063
R^2	0.99925	0.99701	0.99613

B) Michaelis-Menten $y = V_{max}\cdot X/(K_M+X)$

	GS-(*S*)-VG	GS-(*R*)-VG	GS-(*R*,*S*)-VG
V_{max} [mM/min]	0.029 ± 0.0006	0.031 ± 0.0008	0.032 ± 0.0007
K_{cat} [1/s]	63.52 ± 1.31	67.90 ± 1.75	70.09 ± 1.53
K_M [mM]	1.037 ± 0.057	0.828 ± 0.059	0.820 ± 0.041
R^2	0.99572	0.99198	0.99569

Figure 3.25 Determination of kinetic parameters for LigG-TD in the cleavage of GS-VG using the Hill (**A**) and Michaelis-Menten (**B**) equation. For determination of the kinetic parameters, 1 mL reactions were performed using 0.2 mM and 5 mM GS-VG (section 2.3.6.1), the enzyme reaction rate was plotted against the substrate concentration.

This increased substrate concentration, however, led to new problems, which were not observed in the test experiments using only 0.8 mM GS-VG. Due to the strong absorbance of 4 mM GS-VG at 340 nm, the wavelength for detection of NADPH consumption, no NADPH decrease was detectable. Therefore, the wavelength to measure the NADPH decrease was shifted to 360 nm, as this wavelength is better suited to detect NADPH in the presents of GS-VG (Figure 3.26).

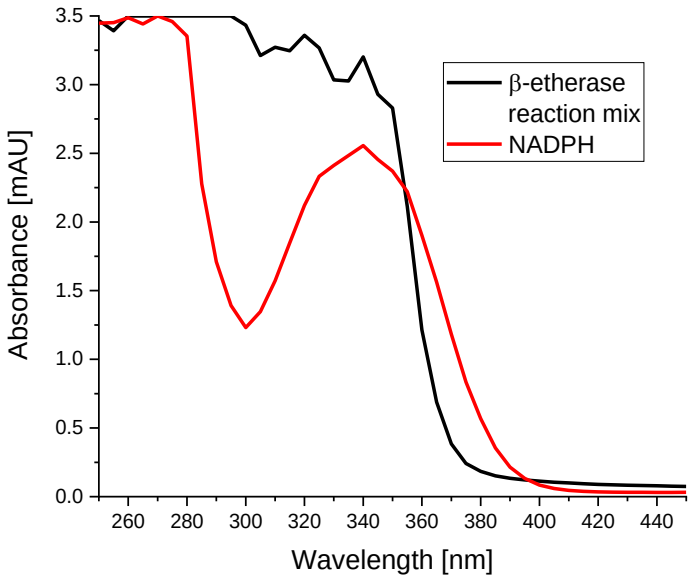


Figure 3.26 Absorbance spectra of 1.2 mM NADPH and the β-etherase reaction mixture containing 4 mM racemic GS-VG.

The final conditions used for the glutathione reductase based HTS assay are summarized in Table 3.7.

Table 3.7 Final reaction conditions for the glutathione reductase assay.

	Component	**End concentration**
100 µL β-etherase reaction mix	8 mM GS-VG 100 mM Tris/HCl pH 8.5 10 mM GSH	4 mM GS-VG 50 mM Tris/HCl pH 8.5 5 mM GSH
50 µL buffer	200 mM Tris/HCl pH 8.5 20 mM GSH	50 mM Tris/HCl pH 8.5 5 mM GSH
25 µL NADPH + glutathione reductase	9.6 mM NADPH 1/100 glutathione reductase	1.2 mM NADPH 0.25 µL glutathione reductase
25 µL cell lysate	Cell lysate 1/10 dilute	
Measure absorbance at 360 nm over 20 min		

To check the accuracy of the glutathione reductase based HTS assay, a microtiter plate containing cell lysate of *E. coli* BL21 (DE3) Gold cells expressing LigG-TD wild-type was screened and the NADPH decrease measured in each well. Based on this, a variation coefficient (standard deviation normalized to the mean value) of this wild-type plate for NADPH consumption was calculated with 12.7%.

3.4.4 Site-directed mutagenesis of the LigG-TD active site

To increase the activity of LigG-TD in the cleavage of GS-VG by protein, both enantiomers of the substrate were first docked in the crystal structure of LigG-TD (section 3.4.2) and a high-throughput activity assay was developed (section 3.4.3). Based on the docking result, residues were selected for a first round of mutagenesis. In previous mutagenesis experiments of Picart et al.,[59] which were herein repeated for a better comparison, the amino acid positions V11, V108, M116 and F165 were mutated. Complementary to these, additional mutations at amino acid positions C12, C15, A55, D104, Y112, and N223 were introduced. Most of these residues were chosen based on direct interactions with substrates GSH and GS-VG, as revealed by substrate docking. In this first round of mutagenesis, most of the selected amino acids were replaced by alanine to gain initial information about their impact on enzyme activity.

Except of A55S, all other amino acid exchanges resulted in a reduced LigG-TD activity (Figure 3.27). Mutations C12S as well as all tested Y112 mutations inactivated the enzyme completely. Hence, C12 and Y112 seem to be essential for the LigG-TD catalysis. In contrast, the exchange of A55 to serine increased the enzyme activity by about 20%. Since every mutation had a serious effect on the activity of LigG-TD, the docking of the substrate seems to be a good description for the binding of the substrates or at least the binding area. Serval of the mutated amino acids were selected for subsequent site-saturation mutagenesis

Glutathione lyase	Enzyme amount [µg]	Specific activity [U/mg]	Rate Mutant/WT
LigG	0.5	47.50	
LigG-TD	0.5	12.02	1.00
LigG-TD*	5	11.98	1.00
LigG-TD V11G	0.5	3.90	0.32
LigG-TD C12S	10	N. A.	N.A.
LigG-TD C15S	0.5	1.32	0.11
LigG-TD A55S	0.5	14.90	1.24
LigG-TD A55S*	5	14.30	1.19
LigG-TD D104A	0.5	11.10	0.92
LigG-TD D104E	0.5	5.96	0.50
LigG-TD R105A	0.5	3.68	0.31
LigG-TD Y112A	10	N.A.	-
LigG-TD Y112F	10	0.02	0.00
LigG-TD Y112W	10	N.A.	-
LigG-TD V108Y Y112A	10	0.01	0.00
LigG-TD V108A	0.5	3.48	0.29
LigG-TD M116A	0.5	0.48	0.04
LigG-TD F165A	0.5	5.56	0.46
LigG-TD F165M	0.5	4.68	0.39
LigG-TD N223A	0.5	1.08	0.09

* = substrate concentration of 4 mM GS-VG

N.A. = not active

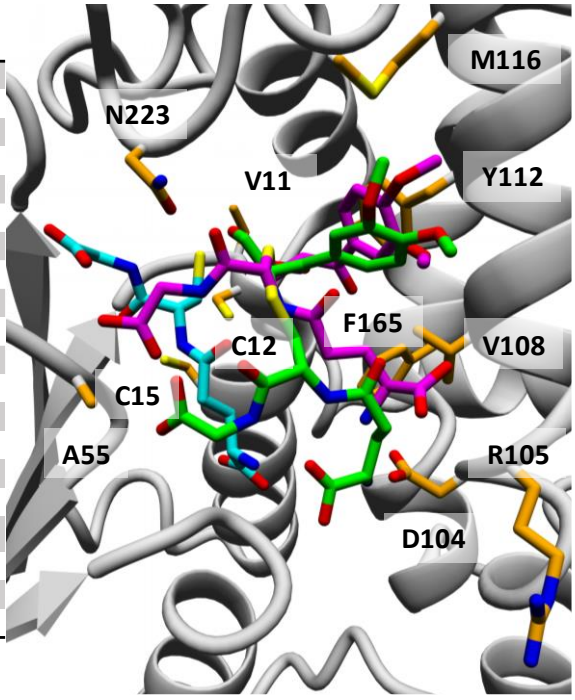


Figure 3.27 Left: Specific activities of LigG-TD wild-type and single mutants with amino acid exchanges around the active site. The reactions were performed in 1 mL scale with 0.8 mM or 4 mM (marked with *) GS-VG. Right:

Active site of LigG-TD co-crystallized with GSH (turquoise) with both enantiomers of GS-VG (GS-*(R)*-VG in magenta and GS-(*S*)-VG in green) docked in. The amino acids that were selected for mutagenesis are highlighted in orange.

3.4.5 Site-saturation mutagenesis of LigG-TD

Site-saturation mutagenesis of residues V11, A55, D104, V108, M116, F165 and N223 was performed using degenerated codons. The desired codon combination NDT, VMA, ATG and TGG was successfully introduced in libraries V11X, A55X, D104X and F165X using Q5-mutagenesis. For positions V108, M116 and N223, however, the Q5-mutagenesis approach was not successful. Therefore, for these positions NNS libraries were generated using the Quikchange protocol.

To investigate the quality of the SSM libraries, the isolated plasmid libraries were sequenced. Figure 3.28 shows a sequencing example for the NNS library at position N223X. This confirms the desired randomization of the library.

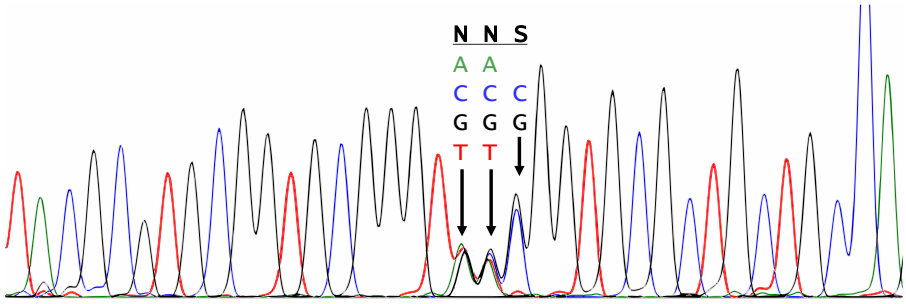


Figure 3.28 Sanger-sequencing of pET28 LigG-TD N223 NNS with the denaturated NNS codon marked.

After successful generation of the mentioned LigG-TD libraries, the glutathione reductase based HTS assay was used for their screening in order to identify mutants with higher activity. Thus, the library clones were grown in microtiter plates for enzyme production and cells were lysed with B-PER. After addition of GS-VG as substrate, NADPH, and glutathione reductase, the reaction was monitored by measuring the absorbance at 360 nm. The slope of the NADPH decrease was used as a reference for the relative activity of the mutants. The slope was normalized to OD_{600}, which was measured after the expression. For better comparison the differences of the single wells to mean value of the wild-type controls were calculated. An example for the screening of mutant library LigG-TD D104X is shown in Table 3.8. The results of the other mutant libraries are shown in section 6.5.

Table 3.8 Results of the LigG-TD D104X library screening using the glutathione reductase based HTS assay. The mutants were expressed in LB media in microtiter plates. The cell pellets were lysed with B-PER and the assay was performed with GS-VG as substrate and NADPH as well as glutathione reductase for detection of the glutathione lyase reaction. The decrease in NADPH concentration was monitored at 360 nm. The slope was determined for each well and normalized to the OD_{600} of the well before cell lysis. Then, the difference of this value to the mean value of the wild-type controls was calculated in percent (section 2.3.7.4). Wells without any cell growth are marked with NG (no growth). Wild-type controls given as red number, empty vector controls as green numbers and mutants chosen for sequencing are highlighted in bold.

	1	2	3	4	5	6	7	8	9	10	11	12
A	**15**	**-54**	39	34	24	27	33	35	**76**	5	-58	36
B	-41	-74	-7	-12	-14	-43	-36	-28	-78	-33	-35	-54
C	-54	-80	**-40**	**-30**	-28	-62	-62	-68	-78	-51	-3	3
D	-81	NG	47	NG	31	16	18	26	54	2	**64**	**60**
E	-16	8	6	-38	26	**-4**	**NG**	-18	**57**	45	**73**	-46
F	-54	-1	-21	17	-16	2	-84	**56**	**49**	**-74**	40	13
G	3	15	36	23	31	-29	5	16	49	44	25	-18
H	-17	35	28	24	26	-86	-60	-23	-56	**73**	**-21**	**-75**

Mutants for which a high positive difference to the mean value of the wild-type control was obtained, were chosen for sequencing. The sequencing results as well as the relative activities in the glutathione reductase based HTS assay of all libraries are listed in Table 3.9.

Table 3.9 Relative activities in the glutathione reductase based HTS assay and corresponding mutations of promising mutants of all screened LigG-TD libraries

Well	Rel. activity	Mutation	Well	Rel. activity	Mutation
V11X (NDT+VMA+ATG+TGG)			V108X (NNS) No. 1		
B5	+61%	V11M	D11	+37%	WT
B8	+105%	WT	G1	+38%	WT
B12	+109%	WT	G3	+36%	WT
C9	+129%	V11A	G7	+27%	WT
C10	+105%	V11I	V108X (NNS) No. 2		
C11	+85%	V11P	C12	+25%	WT
A55X (NDT+VMA+ATG+TGG)			E6	+28%	WT
E5	+114%	A55C	E10	+25%	WT
F8	+28%	A55Q	F12	+33%	WT
G9	+32%	WT	M116X (NNS) No. 1		
H7	+39%	WT	E10	+27%	WT
H9	+34%	A55F	M116X (NNS) No. 2		
D104X (NDT+VMA+ATG+TGG)			E10	+27%	WT
A9	+76%	D104N	F3	+30%	WT
D11	+64%	WT	F10	+25%	WT
D12	+60%	WT	F12	+25%	WT
E9	+57%	WT	N223X (NNS) No. 1		
E11	+73%	D104E	D1	+28%	WT
F8	+56%	D104V	D12	+27%	WT
H10	+73%	D104A	G8	+71%	WT

Well	Rel. activity	Mutation	Well	Rel. activity	Mutation
F165X (NDT+VMA+ATG+TGG) No. 1			N223X (NNS) No. 2		
B3	+65%	WT	B12	+28%	WT
B7	+70%	WT	E12	+34%	WT
B8	+53%	WT	F11	+30%	WT
B12	+62%	WT	H2	+30%	WT
E5	+111%	F165T			
F165X (NDT+VMA+ATG+TGG) No. 2					
B3	+49%	WT			
C6	+37%	WT			

WT = wild-type

Using the glutathione reductase based HTS assay, several LigG-TD mutants with increased relative activity were identified in the site-saturation mutagenesis library screening. Despite this, a high number of sequenced mutants turned out to be LigG-TD wild-type. In libraries V108X, M116X, and N223X, exclusively wild-type enzymes were obtained. Also, on DNA-level exclusively wild-type sequences were observed, which shows a rather high wild-type content in these libraries, probably based on an insufficient DpnI digest.

Since HTS assays usually tend to be quite inaccurately, the activity of the identified mutants of LigG-TD with higher relative activity than wild-type in the library screening has to be evaluated with other more accurate activity assays. Therefore the identified mutants were produced in 100 mL culture, purified and their specific activities in the conversion of GS-VG were determined by HPLC (Table 3.10). For mutants that displayed a higher specific activity than LigG-TD wild-type with the racemic substrate, also the specific activities for conversion of GS-(*R*)-VG and GS-(*S*)-VG were determined.

Table 3.10 Specific activities of selected LigG-TD mutants from the library screening in the conversion of *rac*-GS-VG as well as GS-(*R*)-VG or GS-(*S*)-VG. The reactions were performed in 1 mL scale with 4 mM *rac*-GS-VG or 2 mM GS-(*R*)-VG or GS-(*S*)-VG and analyzed by HPLC. LigG-TD WT, A55F and A55Q were tested with 0.8 mM instead of 4 mM substrate concentration (marked with *).

Mutant	Specific activity for *rac*-GS-VG	Rate Mutant/WT	Specific activity for GS-(*R*)-VG	Specific activity for GS-(*S*)-VG	Rate R/S
WT	11.98 $U \cdot mg^{-1}$	1	13.16 $U \cdot mg^{-1}$	1.30 $U \cdot mg^{-1}$	10.12
WT*	12.02 $U \cdot mg^{-1}$	1	-	-	-
V11A	8.15 $U \cdot mg^{-1}$	0.68	-	-	-
V11I	5.15 $U \cdot mg^{-1}$	0.43	-	-	-
V11M	13.13 $U \cdot mg^{-1}$	1.10	13.69 $U \cdot mg^{-1}$	1.77 $U \cdot mg^{-1}$	7.73
V11P	0.8 $U \cdot mg^{-1}$	0.07	-	-	-
A55C	9.48 $U \cdot mg^{-1}$	0.79	-	-	-
A55F*	8.47 $U \cdot mg^{-1}$	0.70	-	-	-
A55Q*	8.27 $U \cdot mg^{-1}$	0.69	-	-	-

Mutant	Specific activity for *rac*-GS-VG	Rate Mutant/WT	Specific activity for GS-(*R*)-VG	Specific activity for GS-(*S*)-VG	Rate R/S
A55S	14.30 U·mg^{-1}	1.19	17.00 U·mg^{-1}	1.16 U·mg^{-1}	14.66
D104A	10.72 U·mg^{-1}	0.89	-	-	-
D104E	8.50 U·mg^{-1}	0.71	-	-	-
D104N	14.66 U·mg^{-1}	1.22	14.31 U·mg^{-1}	1.19 U·mg^{-1}	12.03
D104V	0.77 U·mg^{-1}	0.06	-	-	-
F165T	9.75 U·mg^{-1}	0.81	-	-	-

As a result the mutants V11M, A55S (identified in the site-directed mutagenesis experiments described in section 3.4.4) and D104N displayed a 10, 19 and 22%, respectively, higher specific activity than LigG-TD wild-type. For mutants A55S and D104N, this increase results in an increase in the conversion of GS-(*R*)-VG. Only for mutant V11M an increased activity towards GS-(*S*)-VG was observed. Therefore, the selectivity of LigG-TD A55S and D104N is increased compared to the wild-type enzyme, whereas it is recued in the case of the V11M mutant.

3.5 Characterization of new glutathione lyases

3.5.1 Expression and characterization of GST3 and LigG817

Beside the search for new β-etherases (section 3.3), also new glutathione lyases were desired. Before Ohta et al. identified GST3 in *Novosphingobium* sp. MBES04, only glutathione lyases of the Omega-class of GSTs were known displaying (*R*)-selectivity in the cleavage of glutathione adducts.[63,68] Later, Kontur et al. described novel glutathione lyases belonging to the Nu-class of GSTs, of which also GST3 is a member.[60] To establish this enzyme family in the group, synthetic genes encoding GST3 and a homolog found in *Novosphingobium* sp. PP1Y were ordered. This homolog was named LigG817 (Genbank accession number WP_041558817) and is 82% identical to GST3 on protein level.

The synthetic genes of GST3 and LigG817 were successfully cloned into pET28a, as confirmed by sequencing, and transformed into chemically competent *E. coli* BL21 (DE3) Gold cells (sections 2.3.4.1, 2.3.4.2, 2.3.4.3, and 2.3.4.4). The subsequent expression and purification of both enzymes were performed according to improved protocols for β-etherases and glutathione lyases (sections 2.3.5.3, 2.3.5.4, and 3.2), and yielded high amounts of soluble and pure enzymes. The yields of purified proteins per litre media were 306 mg·L^{-1} and 517 mg·L^{-1} for GST3 and LigG817, respectively. The ÄKTA chromatograms and the SDS-PAGE analyses of

the purifications of GST3 and LigG817 are comparable to the results for the other β-etherases and glutathione lyases (section 3.2) and are therefore not shown.

For a biochemical characterization of GST3 and LigG817, their melting temperatures and specific activities towards *rac*-GS-VG, GS-(*R*)-VG, and GS-(*S*)-VG were determined and compared to LigG-TD (Table 3.11).

Table 3.11 Melting temperatures and specific activities of GST3 and LigG817 in the conversion of *rac*-GS-VG, GS-(*R*)-VG and GS-(*S*)-VG as compared to LigG-TD and LigG. The reactions were performed in 1 mL scale with 0.8 mM *rac*-GS-VG or 0.4 mM GS-(*R*)-VG or GS-(*S*)-VG (sections 2.3.6.1 and 2.3.6.2). The melting temperatures were determined with the thermofluor assay (section 2.3.6.4). In case of LigG the literature (*R*)/(*S*) by Picart et al.[59] is shown for a better comparison.

	LigG	**LigG-TD**	**GST3**	**LigG817**
Melting temperature	53.5 °C	61.8 °C	48.5 °C	42 °C
Specific activity towards *rac*-GS-VG	47.50 U·mg^{-1}	12.02 U·mg^{-1}	44.97 U·mg^{-1}	29.52 U·mg^{-1}
Specific activity towards GS-(*R*)-VG	-	19.22 U·mg^{-1}	5.59 U·mg^{-1}	2.89 U·mg^{-1}
Specific activity towards GS-(*S*)-VG	-	1.34 U·mg^{-1}	49.43 U·mg^{-1}	35.47 U·mg^{-1}
Rate (*R*)/(*S*)	(18,489.00)[59]	14.31	0.11	0.08
Rate (*S*)/(*R*)	-	0.07	8.84	12.27

Based on the obtained data, GST3 and LigG817 are highly active enzymes. The specific activities of both enzymes in the conversion of *rac*-GS-VG are nearly as high as the one of LigG, the highest active Omega-class glutathione lyase.[59] Especially in the conversion of the GS-(*S*)-VG enantiomer both enzymes reach very high reactions rates. Also, the stereoselectivity of both enzymes is lower than in the case of LigG-TD, the least stereoselective Omega-class glutathione lyase,[59] indicated by the (*R*)/(*S*) and (*S*)/(*R*) rates. When comparing GST3 and LigG817 as Nu-class glutathione lyases, GST3 (first described by Ohta et al.[63,68]) displayed higher specific activities and temperature stability than LigG817.

GST3 and Nu-class glutathione lyases in general seem to be suitable for applications in lignin degradation systems, since Nu-class glutathione lyases perform the best combination of high activity and low selectivity. This was also shown in the investigations of Ohta et al. and Kontur et al.[32,63] Hence, GST3 and other Nu-class glutathione lyases will be interesting candidates for applications in lignin valorization.

3.5.2 Phylogenetic analysis of glutathione lyases

To identify further glutathione lyases homologues, a homology search with subsequent phylogenetic analysis of the obtained sequences was performed. First, BLASTP searches were performed in the nr database of GenBank (release 231) using all known glutathione lyase sequences as queries (BLASTP size 500 sequences, query GenBank accession numbers are listed in section 6.4). Based on the obtained sequences, a multiple sequence alignment was generated and used to construct a phylogenetic tree based on average linkage (UPGMA) using the MAFFT server (section 2.3.3.3). A radial representation of the resulting phylogenetic tree is shown in Figure 3.29.

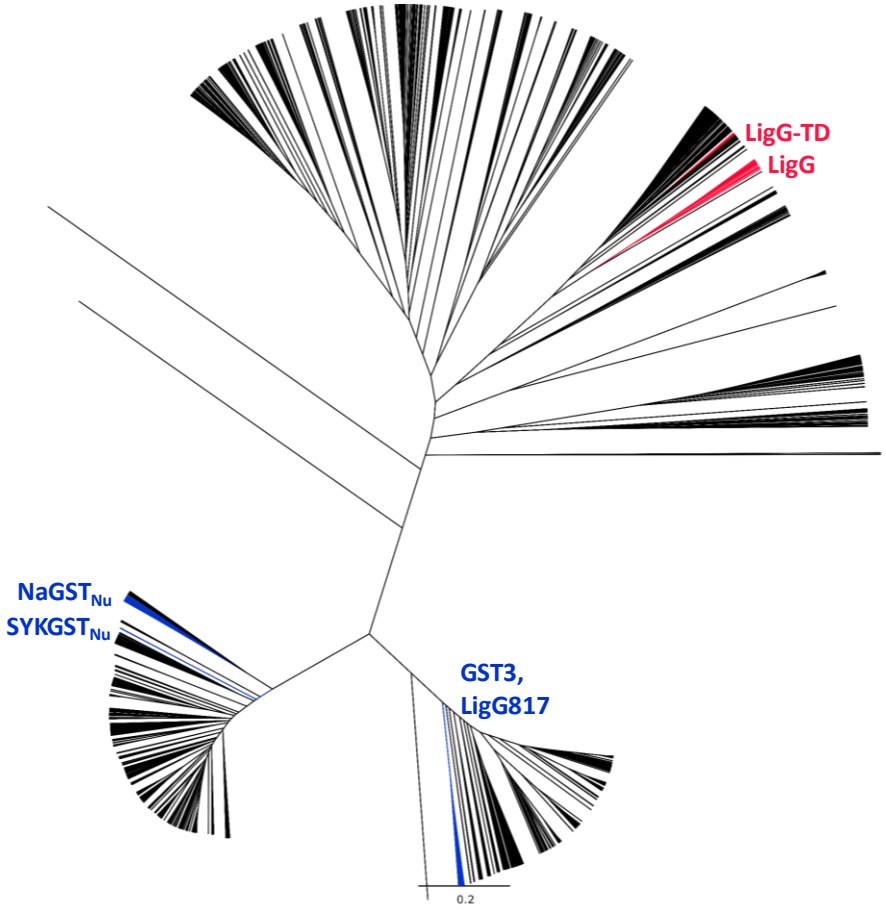


Figure 3.29 Phylogenetic tree of known glutathione lyases and their homologues. Branches with Omega-class glutathione lyases are colored in red, branches with Nu-class enzymes in blue. The multiple sequence alignment and the phylogenetic tree were created with the MAFFT sever (section 2.3.3.3).[97] The sequence database consisted of 2,625 sequences and was constructed by BLASTP searches in the nr database of GenBank (release 231) using all previously known glutathione lyases as queries with a limit of 500 hits per query (GenBank accession numbers of the queries are listed in section 6.4).

The phylogenetic tree reveals a clear evolutionary separation between the Omega-class and Nu-class glutathione lyases. But the phylogenetically diversity is not restricted to the different

groups of glutathione lyases, also within the two groups the enzymes seems to be more diverse compared to the different groups of β-etherases. The members of different types of β-etherases, LigE-, LigF-, and NaLigF-2-type as well as hetero-dimeric β-etherases, are located very dense on the respective branch of the phylogenetic tree (Figure 3.11). In contrast to that, Nu-class enzymes themselves are found on two clearly different branches of the phylogenetic tree. Moreover, $NaGST_{Nu}$ and $SYKGST_{Nu}$ are located on different subbranches. Therefore, the phylogenetic, structural, and functional classification of glutathione lyases, especially Nu-class enzymes, is still in progress and one task for the future.

Beside the general phylogenetic relationships of the different groups of glutathione lyases, a closer look to the branches with known glutathione lyases reveals some areas within the tree, which carry homologs of the known enzymes (Figure 3.30 and Figure 3.31).

The Omega-class enzymes LigG, LigG-NS and GST6 are on one separate branch with 38 other enzymes which could share the same function (Figure 3.30, magenta highlighted part of the tree). As in the case of the β-etherases, additionally to the phylogenetic analysis a PPR analysis was performed (section 3.3.1). The database was created using BLASTP with LigG, LigG-NS, LigG-TD, and GST6 as sequence queries (BLASTP size 500 sequences, nr database, release 231, 1,057 enzymes, GenBank accession numbers of the query sequences and the resulting Omega-class PPR group are listed in section 6.4). The PPR analysis also revealed the phylogenetic and functional relationship of LigG, LigG-NS, and GST6 as it is observed in the phylogenetic tree. Within one PPR-group these three enzymes were grouped together with 38 others. These are the same 38 enzymes found on the same branch of the phylogenetic tree together with LigG, LigG-NS, and GST6 (Figure 3.30, magenta highlighted part of the tree). LigG-TD is phylogenetically separated from the other Omega-class enzymes and also not part of the PPR group of Omega-Class enzymes (Figure 3.30).

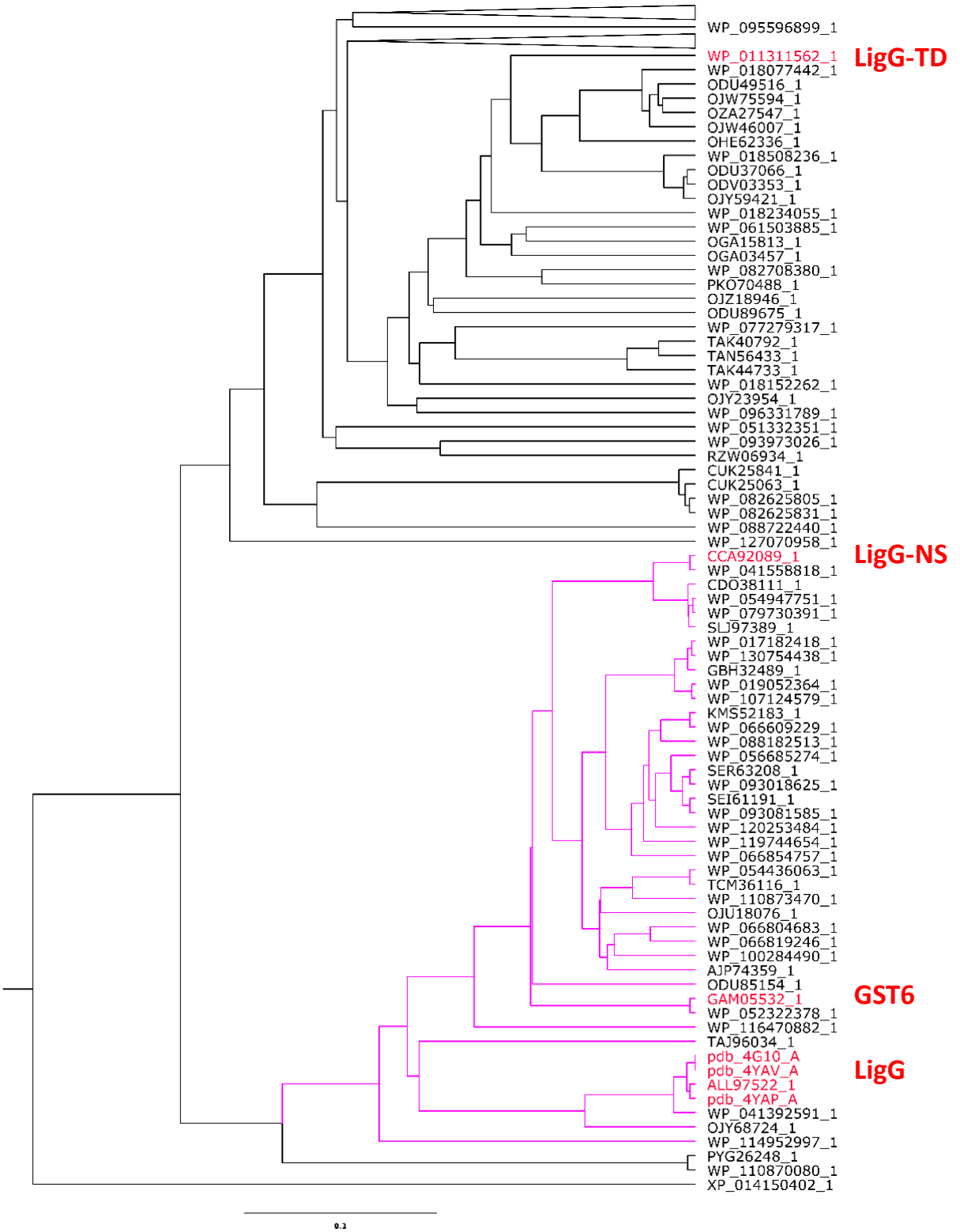


Figure 3.30 Expansion of the phylogenetic tree of known glutathione lyases and their homologs (Figure 3.29). Shown are parts of the phylogenetic tree of known Omega-class glutathione lyases. The sequence database consisted of 2,625 sequences and was constructed by BLASTP searches in the nr database of GenBank (release 231) using all previously known glutathione lyases as queries with a limit of 500 hits per query (GenBank accession numbers of the queries are listed in section 6.4; sections 2.3.3.1 and 2.3.3.2). The GenBank accession numbers of known Omega-class glutathione lyases are highlighted in red, whereas the branch of the tree including the sequences grouped by the PPR analysis is colored in magenta.

Also, in the case of the Nu-class enzymes, beside the phylogenetic analysis the PPR approach was used to group these enzymes. The database was created using BLASTP with LigG817, GST3, $NaGST_{Nu}$, and $SYKGST_{Nu}$ as sequence queries (BLASTP size 500 sequences, nr database, release 231, 1,567 enzymes, GenBank accession numbers of the query sequences are listed in section 6.4). In contrast to the case of the Omega-class glutathione lyases, the PPR analysis was not successful, all enzymes were grouped within two groups. Therefore, the PPR analysis of Nu-class glutathione lyases yielded in no further information.

Nevertheless, a closer look to the phylogenetic tree revealed homologous enzymes, which could be the starting point for further investigations of the Nu-class glutathione lyases. GST3 and LigG817 are located on a subbranch with 4 other enzymes, whereas the subbranch of $SYKGST_{Nu}$ includes of 2 further enzymes (Figure 3.31).

For further studies especially enzymes on the branch with LigG, LigG-NS and GST6 are promising candidates for identification of new Omega-class enzymes. For the Nu-class enzymes on the branches with $SYKGST_{NU}$, GST3 and LigG817 are also enzymes which could be tested for glutathione lyase functionality with good chances for success. One the other hand also more unrelated genes can share the same function and are maybe worth to be tested. A good example for this is LigG-TD phylogenetic separated from the other Omega-class enzymes, but with the same function. Altogether, further investigations are needed to generate a better understanding of the phylogenetic relationships of glutathione lyases, including the separation in different subgroups, and for the identification of novel enzymes with novel biochemical characteristics. The total number of characterized glutathione lyases is still low, so there is still a huge potential to find glutathione lyases, which are more suitable for a desired application, than previously known enzymes.

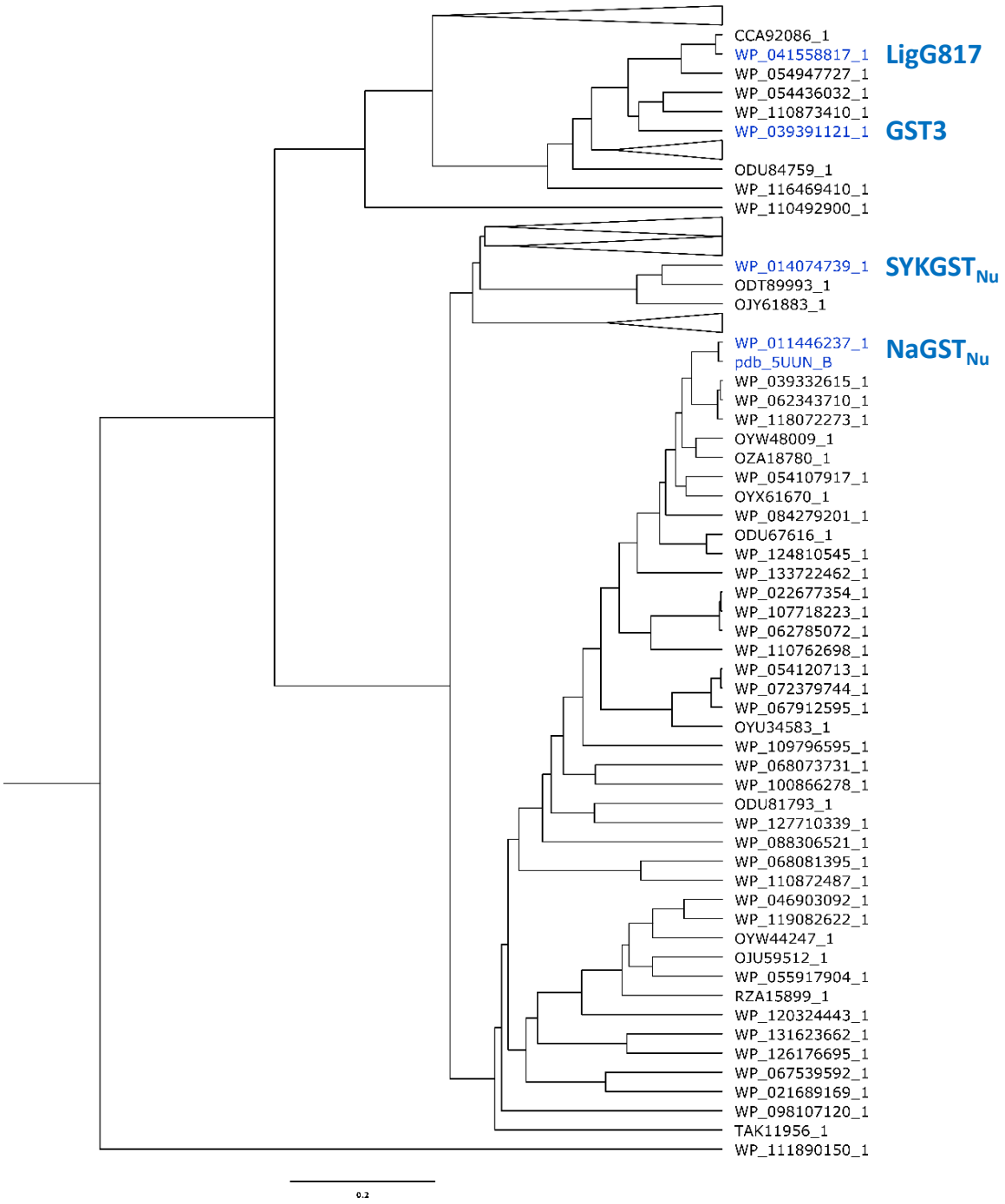


Figure 3.31 Expansion of the phylogenetic tree of known glutathione lyases and their homologs (Figure 3.29). Shown are parts of the phylogenetic tree of known Nu-class glutathione lyases. The sequence database consisted of 2,625 sequences and was constructed by BLASTP searches in the nr database of GenBank (release 231) using all previously known glutathione lyases as queries with a limit of 500 hits per query (GenBank accession numbers of the queries are listed in section 6.4; sections 2.3.3.1 and 2.3.3.2). The GenBank accession numbers of known Nu-class glutathione lyases are highlighted in blue.

3.6 Investigation of thioether lignan cleavage

Some lignans are bioactive compounds. Keto-thioether lignans, structurally related to the keto-ether lignin model compounds, are investigated due to their potential anti-leishmanial

activity.[105] The chemical synthesis of the racemic keto-thioether homologue to the keto-ether lignin model substrates should be uncomplicated. Therefore, β-etherases, known to cleave the keto-ether bound with perfect stereospecificity, should be tested in the chiral resolution of keto-thioether compounds, to produce enantiopure keto-thioether compounds for bioactivity tests (Figure 3.32).

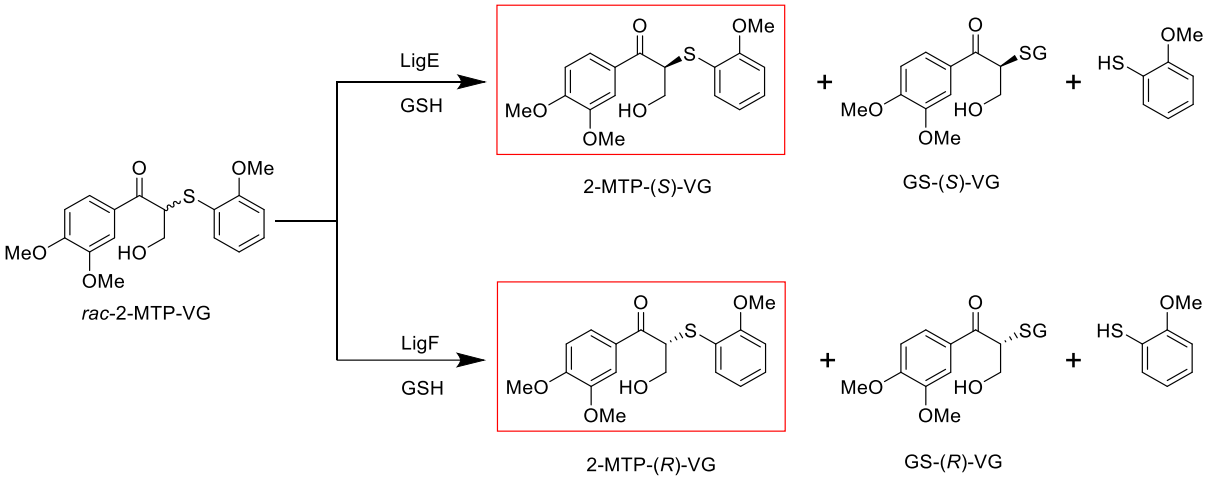


Figure 3.32 Reaction scheme for the chiral resolution of 2-MTP-VG using β-etherases.

Due to the limited number of commercially available methoxy-thiophenol derivates, the system was tested first with 2-methoxy-thiophenol keto-thioether 2-MTP-VG. The synthesis of 2-MTP-VG was successful, but more challenging than accepted. The hydroxy methylation reaction had a rather low product yield of 41.9% and was not as efficient as in the synthesis of 2,6-MP-VG and VN-VG with 84% and 75% product yield, respectively (section 3.1). Nevertheless, the amounts of 2-MTP-VG was sufficient for testing β-etherase activity on keto-thioether compounds.

In the analysis of the cleavage of lignin model compound by β-etherase reactions, often the small phenolic compound released during the reaction is used for HPLC quantification of the conversion. Unfortunately, the corresponding 2-methoxy-thiophenol in the cleavage of 2-MTP-VG seems to be not stable under the used conditions, since with increasing concentrations the peak shape was changing and no correlation between concentration and absorbance in the trails to create calibration curves were found. Also, the second product, GS-VG, is due to its high polarity not easy to quantify, since the compound elutes directly after the injection peak. Therefore, the reaction was followed based on the decrease of the substrate peak.

The cleavage of 2-MTP-VG was tested with the β-etherases LigE, LigE-NA, LigE-NS, LigF, LigF-NA, LigF-NS, and LigP. Since glutathione lyases are also active on keto-thioether (GSH-adducts) LigG, LigG-NS and LigG-TD were used as well in testing the cleavage of 2-MTP-VG. Unfortunately, in most reactions no conversion was observed and in the ones with visible substrates decrease the conversion even at high biocatalyst concentrations (up to 0.5 mg·mL^{-1}) was very low. Highest conversion rates of 21% and 14% were achieved in reactions with LigE and LigF-NA, respectively, based on substrate decrease (Figure 3.33).

Since the quantification was done based on substrate decrease it could be questioned, if real conversion or just variation in the substrate peak integrals due to measuring errors were observed in the reaction with thioether compounds. But in some measurements, as in the case of LigE (Figure 3.33, black and magenta chromatogram), the substrate peak was significantly decreased and also 2-methoxy-thiophenol detected (not quantified but detected), therefore it is very likely that 2-MTP-VG was converted in these reactions.

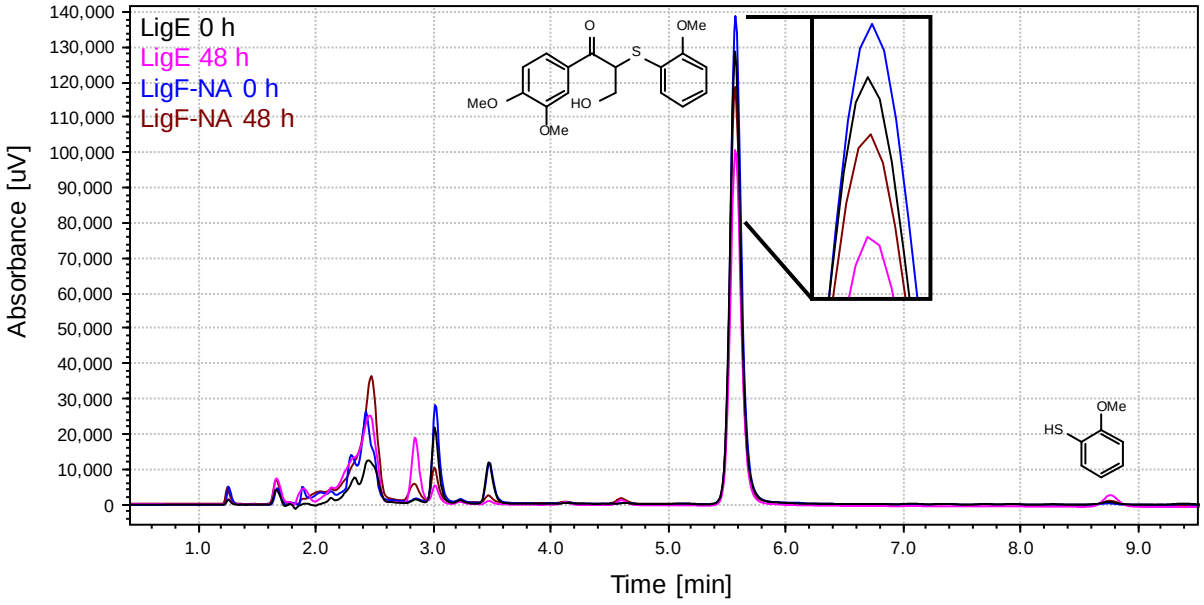


Figure 3.33 HPLC analyses of thioether cleavage of 0.2 mM 2-MTP-VG with LigE and LigF-NA. Shown are the sample measurements after reaction start (LigE black; LigF-NA blue) and after two days of reaction (LigE magenta; LigF-NA brown). The LigE and LigF-NA concentration in the reaction were 0.120 mg·mL^{-1} and 0.233 mg·mL^{-1}, respectively.

DFT calculations performed in the group of Theoretical Chemistry from Prof. Dr. Christoph Jacob at TU Braunschweig using a simplified glutathione molecule were used to investigate why the cleavage of keto-thioether compounds catalyzed by β-etherases and glutathione lyases is so ineffective. The calculations revealed that the ether bond cleavage catalyzed by β-etherases is an exothermic reaction, whereas cleavage of the corresponding thioether bond

would be endothermic (Figure 3.34). This explains the experimental observations that β-etherases are able to cleave lignin model compounds containing β-*O*-4 aryl ether bonds very effectively, whereas in the reactions with 2-MTP-VG no cleavage or low conversion rates were observed.

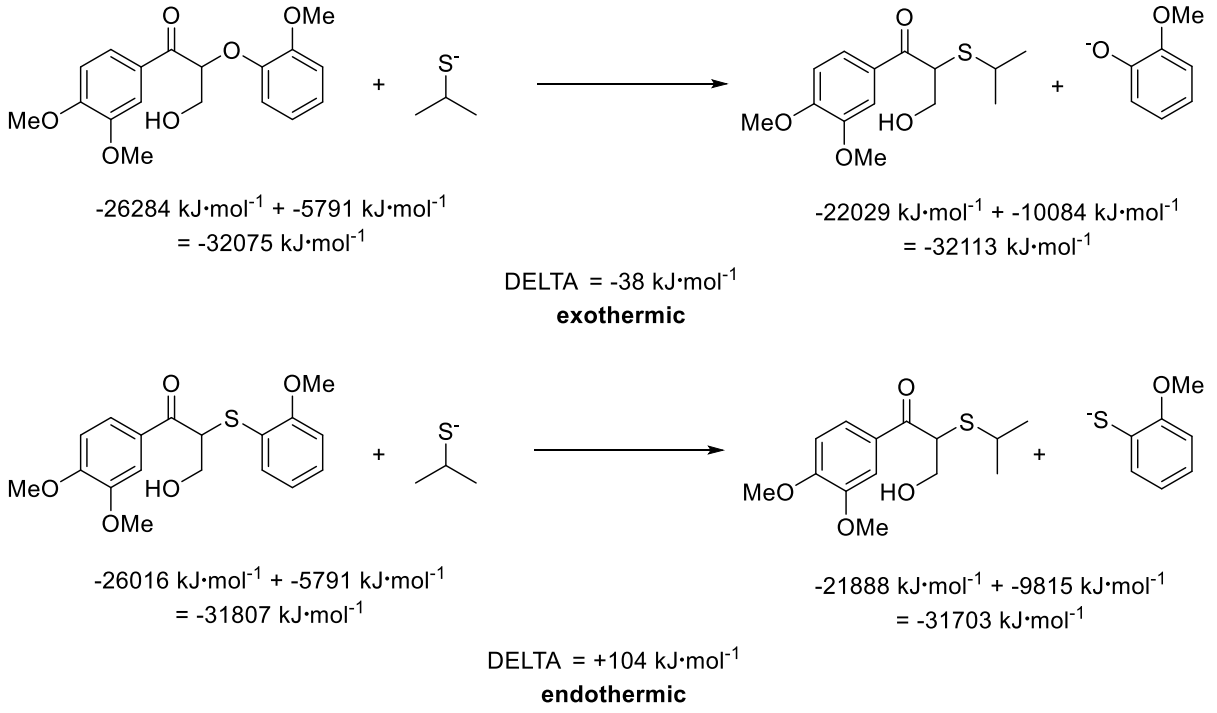


Figure 3.34 DFT calculation results for the ether and thioether cleavage. Shown are the chemical structures of substrates and products with their respective calculated energies. DFT calculations were performed with the program ADF. Water was used as solvent for the calculations with the COSMO model (Conductor like Screening Model). Calculations were performed in the group of Prof. Dr. Christoph Jacob at the Institute for Physical and Theoretical Chemistry of TU Braunschweig.

In summary, the results indicate that β-etherases and glutathione lyases cannot be applied to cleave thioether bond-containing lignans effectively. The low conversion rates as well as the very high enzyme concentrations at low substrate concentration (0.2 mM 2-MTP-VG) are factors that indicate that a preparative thioether cleavage with β-etherases or glutathione lyases is not senseful. In the used chiral resolution approach the chemical properties of the undesired enantiomer are changed, therefore it is critical to reach full conversion of the undesired enantiomer to obtain the desired enantiomer enantiopure. Based on these results this will be very hard to achieve.

3.7 β-Etherase inhibitor synthesis and binding studies

How the substrates are bond in the active center of β-etherases is still unknown. One way to investigate this substrate binding could be co-crystallization experiments with inhibitors. In contrast to the natural substrate, inhibitors are not converted by the enzymes and can be used in the crystallization also together with the co-substrate GSH. In the desired inhibitor the ether bond oxygen atom is replaced by a carbon atom (molecule **12**, Figure 3.35).

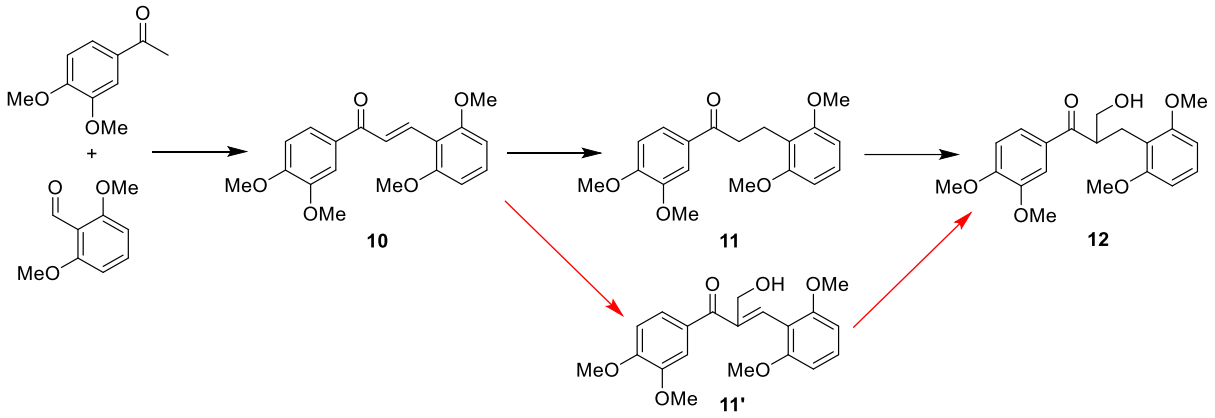


Figure 3.35 Desired and alternative (red arrows) synthesis route for the β-etherase-inhibitor **11** and **12**.

Within the three-step inhibitor synthesis the final hydroxy methylation reaction is the most challenging step, analog to the synthesis of lignin model compounds (3.1). Whereas the molecules **10** and **11** were synthesized successfully, confirmed by NMR and mass spectroscopy, no product formation of compound **12**, with several reaction conditions tested, was observed. Most likely, no reaction occurred based on problems in the deprotonation of compound **11**. For this serval bases were tested (Na_2CO_3, KOH, DIPEA or lithium diisopropylamide). Even with the very strong metal base lithium diisopropylamide no reaction was observed. In the synthesis route of the normal lignin model compounds the α-carbon is additionally activated by the β-ether group. Since also C-C double bonds lead to increased C-H acidity, the hydroxy methylation of compound **10** would be an option for further synthesis attempts.

As **12** was not synthesized successfully, for further tests **11** was used. Since protein crystallization is an expensive and time-consuming method the binding of **11** to β-etherases should be tested first in solution. To investigate this binding MST (micro scale thermophoresis) was chosen. MST is a technique, which quantifies the binding based on the different respond of enzyme and enzyme-ligand-complexes to IR-radiation. Since MST is a method which requires a labelling dye, the results have to be taken with care. Furthermore, the

determination of a K_D value is only possible if a full binding curve with saturation of the binding is measured. In the case of Inhibitor **11** the solubility in a 5% DMSO-water mixture is limited, therefore 1 mM was the highest tested concentration. As β-etherase for testing the binding of **11** LigF-NA was chosen, since the comparison of the crystal structures of LigF and LigF-NA with bound inhibitor would generate important understanding why LigF-NA is more active than LigF.

The binding of **11** to LigF-NA was tested with and without GSH. Without GSH no differences in the MST measurements at the tested concentrations of **11** were observed. When GSH was added small changes in the binding curves at different concentrations of **11** were measured, but no saturation of the binding was observed. Due to this it is not clear, if real binding or only artefacts were measured and no K_D was determined. To generate a clearer result higher concentrations of **11** would be necessary. But since for crystallization normally inhibitor concentration 10 times higher than the K_D are used and the K_D of **11** in the binding to LigF-NA seems to be above 1 mM, it is clear that compound **11** is not suitable for co-crystallization and binding experiments for β-etherases.

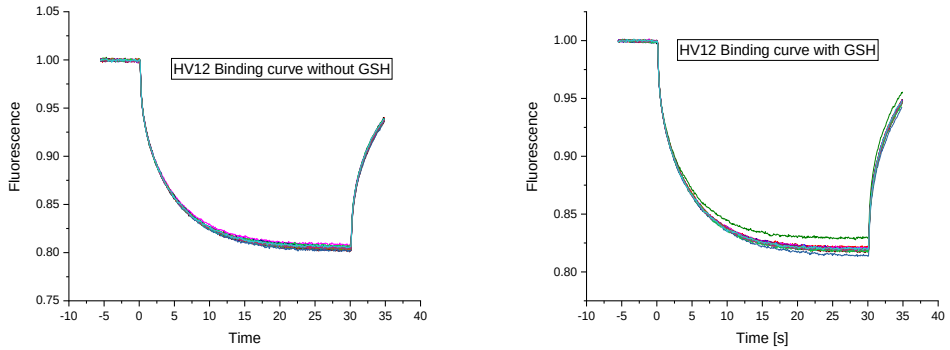


Figure 3.36 Result of the MST binding experiment of LigF-NA and compound **11**. The measurement was performed with 50 nM LigF-NA and concentrations of **11** between 1 and 0 mM, with and without 0.35 mM GSH.

Properly compound **12** would have better solubility and binding properties compared to **11**, due to the hydroxy methyl group, which is also found in the substrates of β-etherases. The higher solubility of **12** is also indicated by the EPA toxicity estimation software tool (TEST, https://comptox.epa.gov/dashboard/predictions/index), which calculated the solubility in water of **11** and **12** with 40 $mg \cdot L^{-1}$ and 73 $mg \cdot L^{-1}$, respectively.

4 Discussion

During this project the limited number of GSH-dependent lignin-degrading enzymes was increased by public database mining using the PPR algorithm. Subsequently, β-etherases and glutathione lyases were analyzed phylogenetically to investigate the relationship of the different groups of GSH-dependent lignin-degrading enzymes. The increased number of β-etherase sequences was used to identify sequence motifs and conserved amino acids as well as to correlate this information with structural data. Regarding the protein engineering of glutathione lyases, the first high-throughput activity assay for these enzymes was developed and successfully applied in the protein engineering of LigG-TD. In the following, the main results of this thesis are discussed in more detail.

4.1 Determined specific activities of β-etherases and glutathione lyases differ from literature data but follow the same trend

Specific activities of known β-etherases and glutathione lyases on defined lignin model compounds have been reported in literature.[30,52,58,59] The activity values determined in this thesis differ, in some cases tremendously, from previous results of Picart et al. (Table 4.1).[30,59] The trend of activities and selectivities, however, is the same.[30,59] This is also confirmed by data of Jana Husarcikova (PhD student at the Institute for Biochemistry, Biotechnology and Bioinformatic in the group of Prof. Dr. Schallmey), who measured again the specific activities of all previously known β-etherases and glutathione lyases reported by Picart et al.[30,59]

Table 4.1 Comparison of determined specific activity values of β-etherases and glutathione lyases with literature data.[30,59]

Enzyme	Substrate	Literature value	Measured value	Difference
LigE	2,6-MP-VG	5.93 U·mg^{-1}	0.65 U·mg^{-1}	-89%
LigF-NA	2,6-MP-VG	6.79 U·mg^{-1}	2.06 U·mg^{-1}	-70%
LigG-TD	GS-(*R*)-VG	36.2 U·mg^{-1}	13.16 U·mg^{-1}	-64%
LigG-TD	GS-(*S*)-VG	0.12·U·mg^{-1}	1.30 U·mg^{-1}	+983%

Despite the huge differences between literature data and the herein measured values, these differences can be partially explained by the use of different enzyme and substrate concentrations. The experiments by Picart et al. were performed with higher enzyme concentrations and in the case of glutathione lyase reactions with much less substrate (4 mM

in this thesis and 0.4 mM by Picart et al.).[30,59] When reaction conditions of Picart et al.[30,59] were used, complete conversion of the substrate was observed in many cases within less than 5 min. Within this short time frame, however, it is difficult to take a sufficient number of samples for the determination of the enzyme activity. Hence, the enzyme concentration was lowered to increase the reaction time. Moreover, the substrate concentration of the glutathione lyase reaction was increased to use a substrate concentration above the K_M value for activity determination (Figure 3.25). Since in addition to the enzyme and substrate concentrations probably also the sample time points (not published by Picart et al.[30,59]) differed, it is not surprising that the exact activity values could not be reproduced. Nevertheless, the trend for the enzyme activities and selectivities were confirmed, which is more important than the absolute values.

The other herein determined catalytic parameters such as selectivity, pH and temperature optima of LigE, LigF-NA, and all newly characterized β-etherases are in perfect agreement with literature data.[30,58,68] The optimal pH for every herein tested enzyme was determined at pH 9, whereas in literature values between pH 9 and 9.5 are described in nearly all cases (only exception is GST5).[30,68] The same holds true for the absolute enantioselectivity of LigE-type and LigF-type β-etherases for (*R*)- and (*S*)-configurated substrates, respectively.[30,58]

4.2 Optimized protocols for the production of β-etherases, glutathione lyases and lignin model compounds lead to increased yields and reduced work effort

The production of (purified) β-etherases and glutathione lyases in larger quantities is very important for their biochemical characterization as well as for a possible application of these enzymes in lignin depolymerization. Using the herein optimized protocols, the yield of purified enzymes was increased 2- to 4-fold in the case of LigE and LigF-NA, respectively, compared to published protocols.[30] At the same time, the amount of work was even reduced. Especially the new β-etherases, which were identified by database mining (section 3.3), can be produced very well using these protocols, as in the case of LigE915 nearly 500 mg purified protein per litre culture could be obtained. If even higher amounts of enzymes are required, e.g. for larger scale applications, the production yield and efficiency could be further optimized using a bioreactor instead of shake flasks as well as high-density cultivation media.[106,107] In such

processes, *E. coli* cell yields above 100 g dry cell weight per litre media are possible, which corresponds to protein yields of serval grams per litre media.[107]

Next to the enzymes, also lignin model compounds are required for the characterization of lignin-degrading enzymes. Therefore, an efficient strategy for the synthesis of those model compounds in appropriate time and scale is required. Within this thesis, two of the three steps towards different β-*O*-4 aryl ether model compound have been improved. First, the efficiency of the bromination reaction, the first step of the synthesis, was optimized. The amount of produced product per reaction was increased 3-fold compared to the protocol published by Picart et al.[30] with a consistent isolated yield of 80%. Gall et al. used a different protocol for the bromination of acetophenone derivates, producing similar amounts of product per reaction but with a lower isolated yield of 59%.[31] In the third step, the hydroxy methylation reaction, the reaction yield could be improved and side reactions avoided by choosing the right reaction solvent for each individual substrate. This way, the yield of 2,6-MP-VG was increased from 51% according to Picart et al.[30] to 84% using the optimized protocol. In case of VN-VG, no product could be obtained based on the protocol of Picart et al.[30], whereas 75% isolated yield were achieved with isopropanol as solvent. This yield is comparable to the 79% yield of VN-VG reported by Gall et al. using 1,4-dioxane as solvent.[31]

4.3 β-Etherases and glutathione lyases are useful for the cleavage of β-*O*-4 aryl ethers but not for the chiral resolution of corresponding thioethers

β-Etherases and glutathione lyases were already tested for selective lignin depolymerization, one potential application of these enzymes.[32,34,63,65] Another possible application is the chiral resolution of ether compounds to produce enantiopure products. Especially β-etherases seem to be well suited for this due to their very high enantioselectivities. Unfortunately, the chiral resolution of the thioether 2-MTP-VG using β-etherases and glutathione lyases was not successful (section 3.6). In most cases no conversion was observed. One of the few exceptions was the reaction with LigE (Figure 3.33) yielding low conversion. A complete conversion of the undesired enantiomer, which is required for chiral resolution, was not achieved in any case even though only 0.2 mM substrate concentration has been used. This substrate concentration would be too low for a useful preparative process. Biocatalytic processes at technical scale ideally with work substrate concentrations in the molar range or at least several grams per litre.[108,109] One example is the DERA aldolase used for the synthesis of the

Atorvastatin side chain, which can handle up to 4.6 M aldehyde substrates.[110] In case of lignin-like substrates, these extreme high substrate concentrations will be hard to reach due to the hydrophobic nature and low solubility of the substrates. For example, Husarcikova et al. published a chiral resolution process for the lignin model compound 2,6-MP-VG using a whole cell catalyst. In this process, already at a substrate concentration of 10 mM with addition of 5% methanol as co-solvent the substrate was not fully dissolved.[111] Therefore, to further increase the substrate concentration the substrate needs to be added as a solid. Such solid-to-solid or solid-to-dissolved conversions with very high substrate concentrations are successfully used for example in the biocatalytically production of peptides.[112]

Regarding the optimization of the enzymatic thioether cleavage, one reason for the low conversion rates is that the reaction is endothermic, therefore it requires energy form the environment. Typical optimization procedures used for endothermic processes would likely not lead to any improvement. Following the principles of Le Chatelier for endothermic reactions, the reaction temperature could be increased to increase conversion rate of the reaction. An increased reaction temperature, however, has two drawbacks, especially since the reaction temperature of 25 °C would have to be increased significantly to yield a significant conversion. First, an increased reaction temperature will have a negative effect on enzyme stability and hence activity. All tested β-etherases (Table 1.1 and Table 3.6) display rather low temperature optima, mostly between 20 and 30 °C.[30,68] Second, the chiral center in the thioether substrate is in α-position to a keto group. It is known that carbonyl compounds with a chiral center on an acidic α-carbon can racemize via an intramolecular keto-enol tautomerism (Figure 4.1).[113] This racemization would be increased as well at higher reaction temperatures, following the Arrhenius equation. Therefore, the options to improve the etherase-catalyzed chiral resolution of thioethers are limited.

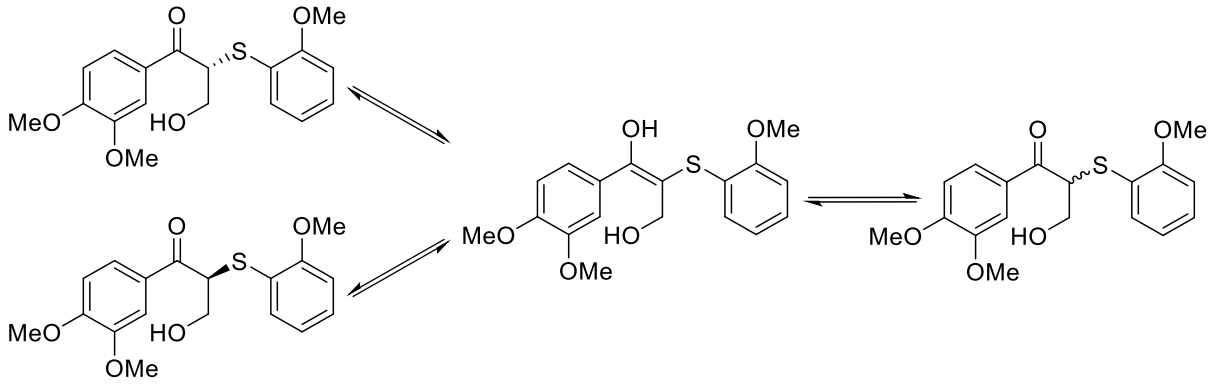


Figure 4.1 Reaction scheme of the racemization of the thioether substrate 2-MTP-VG.

4.4 The PPR algorithm works very successfully for the identification of new lignin-degrading enzymes

Public database mining for the identification of novel enzymes with a desired function is an important field of modern biosciences. The PPR algorithm was described as a powerful and useful approach in this field.[73] In this study, 96 new putative β-etherases have been identified using the PPR algorithm based on sequence datasets that had been generated by homology searches. Though only 13 of the 96 putative enzymes have been investigated herein, confirming their expected β-etherase activity, it can safely be assume that all other 83 putative enzymes will exhibit β-etherase activity, too, considering their placement together with known LigE- and LigF-type enzymes in the phylogenetic tree and their grouping by PPR. Also Busk and Lange, who developed the PPR algorithm, assumed that enzymes grouped together by the PPR algorithm would share the same function, even if only a few members of the group have been characterized.[73,74] Considering the high number of still uncharacterized putative β-etherases, 39 LigE- and 44 LigF-type enzymes, there is still potential for the identification of more active biocatalyst. Furthermore, PPR groups were also identified for NaLigF-2-type β-etherases and Omega-class glutathione lyases. This can be very important to increase the knowledge about these less investigated enzymes.

Additionally, Busk and Lange hypothesized that their PPR algorithm could also identify new enzymes which are phylogenetically unrelated to the previously known ones, while sharing the same activity.[73] In this case, however, only sequences sharing highest sequence homology to known LigE- and LigF-type β-etherases were grouped together with respective known β-etherase sequences during the PPR run. Still, this was quite helpful to distinguish between true LigE-/LigF-homologous β-etherases and other sequence-related glutathione transferases.

The PPR algorithm further enables the automatic identification of conserved sequence patterns in sequences that were grouped together. The same patterns, in principle, can also be inferred manually from a multiple sequence alignment of the same sequence set. Hence, different sequence motifs including strictly conserved amino acids could be obtained for LigE- and LigF-type β-etherases. These include residues S21 in LigE- as well as N13 and S14 in LigF-type enzymes (residue numbering according to LigE and LigF, respectively), which have been suggested to affect β-etherase activity by improved deprotonation of the GSH thiol, a prerequisite for nucleophilic attack at the β-carbon of the β-*O*-4 aryl ether bond.[61,69]

Accordingly, mutagenesis of S21 in LigE and S14 in LigF, as well as mutagenesis of N14 in the hetero-dimeric β-etherase BaeAB from *N. aromaticivorans* (corresponding to N13 in LigF) resulted in significantly reduced to almost complete loss of enzyme activity.[61,69] Beside these three amino acids, several other conserved residues were identified by PPR and the combination of sequence and structural analysis.

Literature concerning conserved sequence motifs in other GSTs are rather rare so far. Allocati et al. described a GXXX[S,T]XXD sequence motif located in the helical domain near the active site of GSTs of several organisms, mammals to bacteria.[114] This motif was not found in LigE- or LigF-type β-etherases, which is not surprising since the helical domain forms the second half of the substrate binding site, considering the large variety of substrates which is converted by different GSTs. In contrast, the GSH-binding seems to be more conserved, which was also observed in the sequence analysis of LigE- and LigF-type enzymes (section 3.3.4). Allocati et al. described further a SXXhh motif, with h being a hydrophobic amino acid, in the GSH-binding site of the GST PmGSTB1-1 from the bacterium *Proteus mirabilis*.[114] Both LigE- (S_{21}XXVW) and LigF-type (S_{14}XXP[L,M]) enzymes are carrying this motif. Furthermore, Widersten et al. described several conserved amino acids in the GSH-binding sites of GSTs.[115] The corresponding amino acids K46, Q53, and E65 are also found in all LigF-type enzymes. Further investigations are required to elucidate the structural and/or functional roles of conserved residues identified in LigE- and LigF-type β-etherases, which were not described in literature so far.

4.5 Phylogenetic analysis of β-etherase- and glutathione lyase-homologs reveals new putative lignin-degrading enzymes

Within this thesis, phylogenetic analyses of β-etherases and glutathione lyases were performed to investigate the relationships of β-etherase and glutathione lyase homologs and to identify new group members. Schallmey et al. used a comparable phylogenetic analysis to group 37 newly identified halohydrin dehalogenases into subtypes.[116] In the phylogenetic analysis, the same LigE-, LigF-, and NaLigF-2-type β-etherase as well as Omega-class glutathione lyases were clustered together as grouped in the PPR analysis. So, the same enzyme groups containing the same sequences were identified with two different approaches, supporting the reliability of the results.

Furthermore, the phylogenetic analysis of β-etherases revealed that LigF-, NaLigF-2-type and hetero-dimeric β-etherases are placed on the same phylogenetic arm carrying also yet unassigned sequences. Hence, all three β-etherase types seem to originate from a common ancestor. Moreover, new types of β-etherases might even be found here and could be interesting for further investigation.

In the case of LigG-TD, Nu-class glutathione lyases and the hetero-dimeric β-etherases, no PPR groups were obtained. In contrast, based on the phylogenetic analyses several homologs of Nu-class glutathione lyases as well as hetero-dimeric β-etherases were identified. Nu-class glutathione lyases and hetero-dimeric β-etherases seems to be quite diverse and the number of direct homologs is rather low (4 for GST3, 2 for $SYKGST_{Nu}$ and 5 for BaeB). This is maybe also the reason why the PPR algorithm was not successful for these groups of enzymes. Investigation of the found homologs of Nu-class glutathione lyases and the hetero-dimeric β etherases will help to increase the knowledge of these relatively new enzyme groups.

Kontur et al. also analyzed the phylogenetic relationship of the GSH-dependent lignin-degrading enzymes.[61] The main results of their phylogenetic analysis are in agreement with the phylogenetic analyses of this thesis. In both cases, the relationship of LigF- and NaLigF-2 enzymes to the hetero-dimeric β-etherases is visible. A new result is the separation of Nu-class glutathione lyases into two phylogenetic subgroups, GST3 homologs and $NaGST_{Nu}$/$SYKGST_{Nu}$ homologs (Figure 3.31). This separation is not visible in the phylogenetic tree published by Kontur et al.[61] The reason for this difference in the phylogenetic trees is probably that Kontur et al. included all GSH-dependent lignin degrading enzymes within one phylogenetic tree, whereas β-etherases and glutathione lyases were analyzed separately in this thesis. Therefore, smaller phylogenetic differences are better visible in the separate phylogenetic analysis.

4.6 Both PPR and phylogenetic analysis have their benefits

With the PPR algorithm and the phylogenetic analysis two different approaches were used for the identification of novel GSH-dependent lignin-degrading enzymes. In all cases, the same putative novel β-etherases and glutathione lyases were found. Comparing both approaches, PPR and phylogenetic analysis both offer individual benefits. Applying PPR on a large sequence dataset including thousands of sequences is an easier and quicker approach than phylogenetic analysis including multiple sequence alignment and tree building of the same dataset,

especially for non-experienced users. Furthermore, the PPR algorithm directly generates information about conserved peptide and sequence regions within the separated group of homologous enzymes. In the other case a separate multiple sequence alignment of the group of homologous enzymes is needed to infer this information.

On the other hand, phylogenetic analysis of a given dataset of homologous sequences does not only allow for grouping sequences into different subgroups and branches of the phylogenetic tree, but also visualizes the phylogenetic relationship of the different subgroups, as in the case of LigF-, NaLigF-2- and BaeAB-type β-etherases.[61,116] This information is lost during the PPR analysis. Furthermore, in the case of Nu-class glutathione lyases and hetero-dimeric β-etherases no PPR group were identified, whereas in the phylogenetic analyses homologs were found, which could be characterized to generate more information on of this enzyme groups. Overall, both approaches PPR and phylogenetic analysis, have their benefits and can be used complementary.

4.7 New indications for lignin-degrading bacteria outside the order *Sphingomonadales*

Beside the knowledge on specific enzymes identified with the PPR algorithm, also new information regarding lignin-degrading microorganisms are generated. Until now, only for bacteria of the families *Sphingomonadaceae* and *Erythrobacteraceae* (both order *Sphingomonadales*) the GSH-dependent degradation of lignin model compounds has been described.[50,61,62,117] Palamuru et al. demonstrated that a bacterial isolate belonging to the family of *Erythrobacteraceae* could grow on the lignin model compound guaiacylglycerol-β-guaiacyl-ether as sole carbon source.[117] Additionally, Cα dehydrogenases, catalyzing the first step in the GSH-dependent β-*O*-4 aryl ether degradation pathway by oxidation of the Cα-hydroxyl group, have been identified in this bacterium.[117] Hence, beside β-etherases, as demonstrated in this study, it is expected that sequences encoding glutathione lyases of the Omega- or Nu-class of GSTs will be present in *Erythrobacter* and *Altererythrobacter* sp., too, to complete the GSH-dependent pathway for degradation of lignin-derived β-*O*-4 aryl ethers.[60,70,71]

Next to the sequences originating from bacteria of the *Sphingomonadales,* LigE- and LigF-type β-etherases were also identified in the marine alphaproteobacterium *Marinicaulis flavus* belonging to the family of *Parvularculaceae* (order *Parvularculales*).[118] Hence, in contrast to

previous assumptions, this is the first indication that enzymes with β-etherase activity can be also found in bacteria outside the order *Sphingomonadales*.[31] Homology searches in the genome of *M. flavus* further revealed the presence of sequences encoding homologs of the Nu-class glutathione lyases and Cα dehydrogenases. Therefore, *M. flavus* seems to harbour a functional GSH-dependent β-*O*-4 aryl ether degradation pathway. In addition, also LigF215 seems to originate from a non-sphingomonad bacterium as the gene was found in a metagenome-assembled genome (MAG) from environmental DNA which has been classified as gammaproteobacterium. This classification, however, should be taken with great care, as it has been shown that such MAGs might contain substantial contamination leading to less reliable phylogenetic binning.[119] Moreover, more than 3% of the respective contig carrying the LigF215 gene contain multiple gaps between 26 to 189 nucleotides, indicating a rather low assembly quality.

Beside the genome of *M. flavus*, also all host organisms of tested β-etherases (Table 3.3) were investigated by homology searches for the other enzymes of the GSH-dependent lignin degradation pathway. In all host organisms Cα dehydrogenases and glutathione lyases homologs were found in the genome, indicating that also these organisms as well harbour the whole GSH-dependent lignin degradation pathway.

4.8 Among the new β-etherases are better biocatalysts than the previously known ones

In 2014, Picart et al. reported LigE from *Sphingobium* sp. SYK-6 and LigF-NA from *N. aromaticivorans* to display highest specific activities among the known β-etherases.[30] None of the later characterized enzymes exhibited higher activity than these two enzymes.[61,66,68] The herein performed characterization of 13 new β-etherases revealed four LigE- and three LigF-type enzymes with up to 3-fold higher activities than LigE and LigF-NA, respectively. Among these are especially the enzymes of *Altererythrobacter* and *Erythrobacter* species (LigE179, LigE283, LigE889, LigF008, LigF921, LigF935). Based on these results, LigE283 from *Altererythrobacter* sp. 66-12 and LigF008 from *Altererythrobacter* sp. Root672 are the new benchmarks for their respective β-etherase-types. Since these enzymes are not only highly active but also highly enantioselective and moderately thermostable, with T_m values above 55 °C, they also constitute excellent biocatalysts.

Also the currently known in vitro systems for lignin depolymerization or whole cell system for the conversion of small molecules could profit from the new β-etherases, since they are all performed with less active enzymes.[34,63–65,111] In 2018, Kontur et al. compared different lignin-degrading bacterial strains, *N. aromaticivorans* DSM 12444, *Novosphingobium* sp.PP1Y, and *S. xenophagum* NBRC 107872, in the metabolism of guaiacylglycerol-β-guaiacyl ether, with the result that *N. aromaticivorans* is the most efficient one.[60] Until now, *N. aromaticivorans* was known as the bacterial strain coding for the most active β-etherases.[30,62] Since the characterized β-etherases are in general less active compared to the other enzymes of the pathway, Cα dehydrogenases and glutathione lyases (see activities of β-etherases and glutathione lyases in Table 1.1, Table 3.5, Table 3.11, and in literature),[30,63,68] β-etherase activity seems to be the bottleneck of the pathway. Therefore, it is very interesting to test the identified *Altererythrobacter* and *Erythrobacter* species containing more active β-etherases.

4.9 The established glutathione reductase based high-throughput activity assay is a good tool for the protein engineering of glutathione lyases

The high-throughput activity assay indirectly measures the activity of the glutathione lyase via online measurement of the NADPH absorbance. It is the first high-throughput assay described for glutathione lyases. The simple setup of the assay without long incubation times and only a few process steps helps to reduce the time and effort required for library screening. Hence, it is possible to screen 10 or more microtiter plates using this glutathione reductase based HTS assay on a normal 8 h working day. Considering a maximum of 10,000 mutants (≈100 microtiter plates) that can be screened in 24 h in microtiter plates under optimal conditions, using for example an automated robotic platform,[120] the efficiency of the developed glutathione lyase activity screening assay is already quite good. Especially compared to the otherwise standardly used HPLC method to determine glutathione lyases activity, the developed HTS assay has by far the higher throughput. Even with advanced chromatographic protocols with multi injection only a few hundred samples per day can be measured by HPLC.[121] Moreover, since these are just measurements of single time points, multiple data points have to be measured to determine enzyme activity accurately.

Beside the high throughput of the assay, the online measurement of the assay reaction is the main benefit. Whereas many other enzymatic assays work discontinuously and require the measurement of several time points, the glutathione lyase assay follows the reaction

continuously in one measurement. On the other hand, the assay also yielded a relatively high number of false positives in the screening of the LigG-TD libraries. One reason for this is the already high activity of the wild-type enzyme in the tested cell lysate. Thus, a small deviation in the amount of glutathione lyase, e.g. based on pipetting inaccuracy, can lead to differences in the measurement. This is also visible in the quite high coefficient of variation which was determined at 12.7% in the screening of a LigG-TD wild-type plate. This means that only LigG-TD variants with significantly increased activity compared to wild type can be identified safely with this assay. Automatization of the screening process would be a possibility to further reduce the error rates of the assay. Beside an increased throughput of microtiter plates that can be screened, the automatization by a robotic platform results in a more precise and reproducible handling of the screening process and therefore, to lower error rates.[122] Also, purification of the glutathione lyase variants in microtiter plates, using for example His MultiTrap plates (GE Healthcare - Solingen, Germany) for IMAC purification, could lead to more precise results. Such a purification would require extra steps and thus, more workload, but would allow to determine the exact enzyme concentration used for the screening. In this work, normalization of each well to the obtained OD_{600} was used as an approximation for the differences in enzyme concentration. The OD_{600}, however, is only sensitive to differences in the cell growth and not to differences in the expression level, which might change due to the introduced mutation. Furthermore, enzyme purification prior to screening would eliminate a possible influence of *E. coli*-innate enzymes that might consume or produce NADPH.

Nevertheless, the herein performed protein engineering of LigG-TD demonstrates that the glutathione lyase assay already works successfully without any further improvements, since several positive LigG-TD were mutants displaying higher activity could be detected.

Likewise, the glutathione lyase assay can also be used for the screening of other glutathione lyases, such as the Nu-class enzymes, as the only requirement is the formation of oxidized glutathione. The Nu-class glutathione lyases are very interesting for future application in lignin depolymerization, as these enzymes are more active and less stereoselective than the Omega-class glutathione lyases. The latter was already reported for GST3, $NaGST_{Nu}$ and $SYKGST_{Nu}$ by Ohta et al. and Kontur et al.,[60,63] and also demonstrated in this thesis with the characterization of LigG817 from *Novosphingobium* sp. PP1Y.

4.10 Docking poses were found that are in agreement with both postulated reaction mechanisms for glutathione lyases

In preparation for the protein engineering of LigG-TD with the aim to increase its activity, the substrate GS-VG was docked into the crystal structure of LigG-TD. For glutathione lyases, two different reaction mechanisms are described in literature (Figure 3.15).[60,71] In the two-step mechanism postulated by Pereira et al. for Omega-class glutathione lyases such as LigG, the GSH-adduct binds to the enzyme first and is cleaved by nucleophilic attack of a cysteine with formation of a covalent GSH-enzyme intermediate.[71] In the other mechanism, GSH and the GSH-adduct both bind at the same time to the active site of the enzyme and GSH directly attacks the GSH-adduct. This mechanism was postulated by Kontur et al.[60] for Nu-class glutathione lyases, and is now used to describe a possible one-step mechanism for Omega-class glutathione lyases (section 3.4.2). During the substrate docking process, substrate binding poses in agreement with both reaction mechanisms were found (Figure 4.2). Therefore, based on the docking alone it is not possible to distinguish which mechanism is valid for LigG-TD, but both mechanisms are plausible.

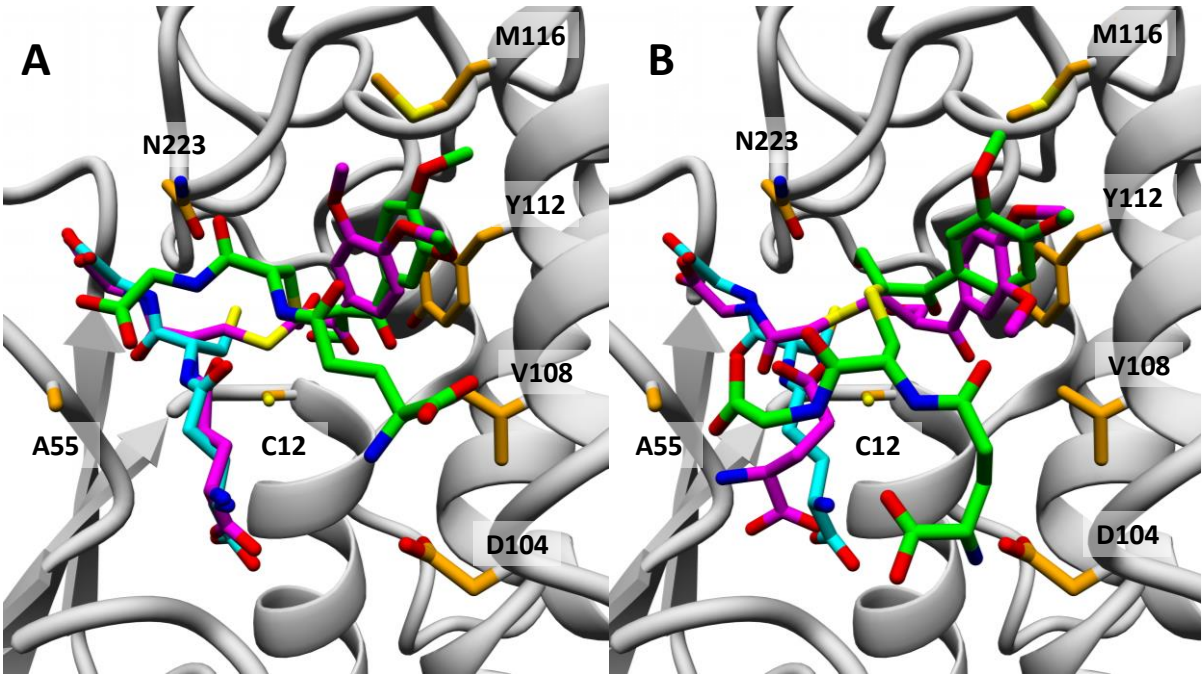


Figure 4.2 Comparison of the LigG-TD substrate docking with the respect of the postulated reaction mechanism. The substrates GS-(*R*)-VG (**A**) and GS-(*S*)-VG (**B**) were docked into the crystal structure of LigG-TD with and without co-crystallized GSH using Vina (implemented in Yasara, section 2.3.3.4). GSH is colored in turquoise, GS-VG docked in apo structure of LigG-TD is shown in magenta, GS-VG docked into the LigG-TD structure co-crystallized with GSH is in green. The docking was performed with 999 docking runs, except for the docking of GS-(*R*)-VG into LigG-TD co-crystallized with GSH (**A**, colored in green). In this case, 200 docking runs were performed. A selection of active site residues chosen for mutagenesis are shown in orange.

Based on possible substrate-enzyme interactions according to the docking results, amino acids of LigG-TD were chosen for mutagenesis through alanine scanning. All amino acid exchanges had a significant impact on LigG-TD activity. Therefore, the obtained docking poses seem to define the substrate binding pocket very well. Comparing the respective docking supporting one or the other postulated reaction mechanism and the resulting interactions between active site residues and the substrate, it is noticed that most interaction partners are the same, independent of the individual docking pose. Exceptions are the residues A55 and D104, which do not directly interact with GS-VG docked into apo-LigG-TD, but as shown in mutation experiments these amino acids have a huge influence to LigG-TD catalysis (see Figure 3.27 and Figure 4.2; how amino acids influence LigG-TD catalysis is discussed in section 4.11). In contrast to that, interactions of both residues with the GS-adduct are visible in the substrate docking according to the mechanism postulated Kontur et al.[60] Nevertheless, since mutations can affect the enzyme activity also indirectly by so called second shell effects, the observations that A55 and D104 do not interact directly with GS-VG docked into apo-LigG-TD can only give indications, but do not disprove the two-step mechanism postulated by Pereira et al.[71] Such an example for a second shell mutation is also found in the protein engineering of LigG-TD for mutagenesis of amino acid position V11 (Figure 4.5). V11 does not directly interact with the substrate, but still influences LigG-TD activity. In conclusion, the docking and the experimental data are in accordance with both mechanisms for glutathione lyases, the one postulated by Pereira et al.[71] and the one according to Kontur et al.[60], and therefore none of the mechanisms can be ruled out for LigG-TD as member the Omega class glutathione lyases.

Considering the number of reaction steps and involved reaction components, both postulated reaction mechanisms differ (Figure 3.15). The mechanism postulated by Kontur et al.[60] for Nu-class glutathione lyases consists of one reaction step with three reaction components, whereas the mechanism postulated by Pereira et al.[71] for Omega-class glutathione lyases consists of two reaction steps with two reaction components each. This will result in different kinetic models to describe the enzyme reaction rate. Therefore, mathematical modelling of the reaction will help to elucidate if the experimentally determined reaction kinetics follow a one- or two-step mechanism.

The formation of GSSG, the second step in the two-step mechanism postulated by Pereira et al., is described as a nucleophilic attack of free GSH on the GSH-enzyme-adduct formed in the first reaction step (see Figure 3.15 **A**).[71] This would require the binding of an additional GSH

molecule. In Omega-class glutathione lyases, no second binding site for GSH in the active site has been described.[60] Nevertheless, docking poses for a second GSH molecule were found in LigG-TD. This second molecule, however, seems to have only very few interactions with LigG-TD (Figure 4.3). On the one hand, these weak interactions could be beneficial for an efficient release of GSSG. On the other hand, for effective GSSG formation stable binding of free GSH in close proximity to the GSH-enzyme-adduct is required. Interestingly, one of the possible interaction partners of the second free GSH molecule is D104, which seems not to interact with the GSH-adduct in the mechanism postulated by Pereira et al.[71]

In general, dockings are just snap shots and do not cover the whole catalysis process. Therefore, it is also not surprising that D104 was not found as an interaction partner in the docking for the first reaction step of the mechanism postulated by Pereira et al.[71], but in the docking for the second reaction step. In contrast to dockings, in silico reaction modelling using QM/MM models allows to follow the whole process and would be beneficial to generate a complete view of the LigG-TD catalysis and to elucidate which reaction mechanism is used by the enzyme.

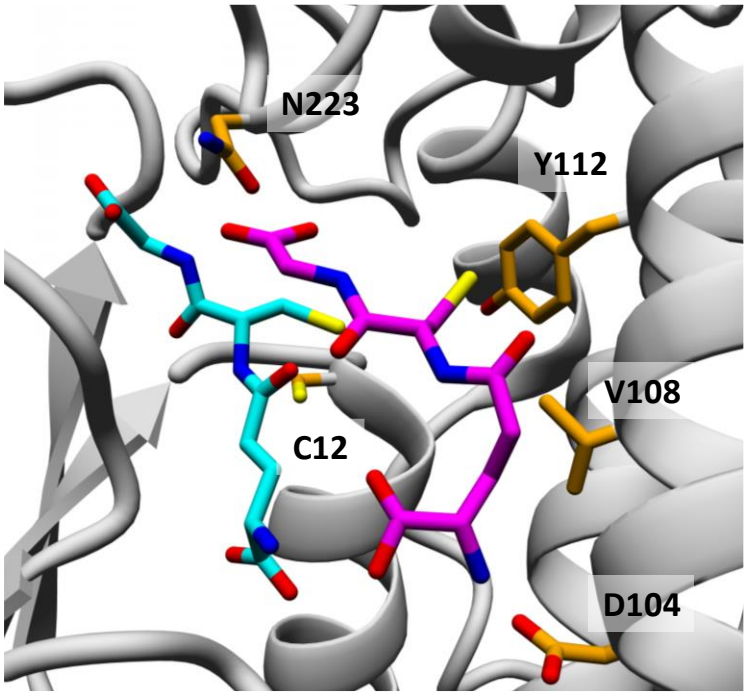


Figure 4.3 Docking of GSH in the crystal structure of LigG-TD, co-crystallized with GSH using AutoDock Vina (implemented in Yasara, section 2.3.3.4). The docked free GSH is shown in magenta, whereas the co-crystallized GSH is shown in turquoise. The docking was performed with 999 docking runs. A selection of amino acids interacting with the docked free GSH molecule (D104, V108, N223) as well as the amino acids C12 and Y112 are highlighted in orange.

Another possibility would be an experimental validation of the two-step mechanism postulated by Pereira et al.[71] by separation of the two catalytic steps. This will be experimentally challenging, since stoichiometric amounts of enzyme are required, and all reaction intermediates have to be stable enough to be separated from the reaction partner. If in the reaction of LigG-TD with GS-VG (Figure 4.4, step 1) in the absence of GSH the product VG could be detected, this would be a strong indication that LigG-TD follows the two-step mechanism postulated by Pereira et al.[71]. Furthermore, this would disprove, in the case of LigG-TD, the one-step mechanism postulated by Kontur et al.[60]. In the best case, the full catalytic cycle could be performed in separate reactions more than ones (see Figure 4.4).

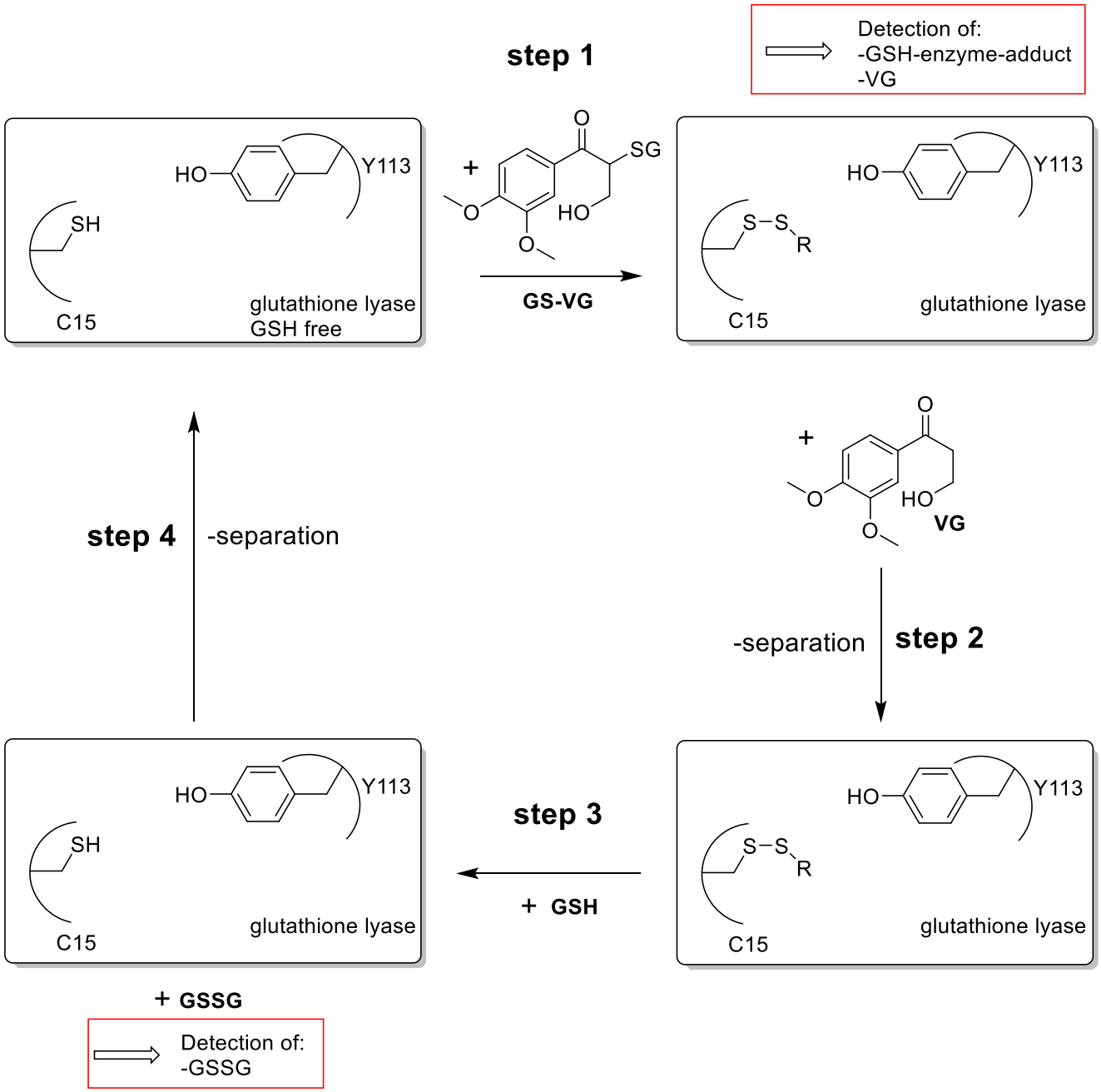


Figure 4.4 Reaction scheme of separate reaction cycle of Omega-class glutathione lyases (LigG numbering) following the two-step mechanism postulated by Pereira et al.[71] The red boxes describe how the reaction steps (step 1 + 3) should be analyzed. In these experiments it is important to use purified GS-VG as substrate for reaction step 1. Usually GS-VG is produced in a β-etherase reaction and the reaction mixture is used without any purification steps as substrate for the glutathione lyase reaction. In this case purified GS-VG has to be used to avoid GSH contamination for the β-etherase reaction.

4.11 Protein engineering of LigG-TD revealed several mutations with positive influence on LigG-TD activity

Several amino acids and amino acid exchanges with a big influence on LigG-TD activity were mentioned in previous sections. A key role in glutathione lyase activity play the amino acids

C15 and Y113 (residue numbering according to LigG), which were already described by Pereira et al.[71] (Figure 3.15). Mutagenesis of one of both amino acids in LigG resulted in complete inactivation of the enzyme.[71] Likewise, exchanging the respective residues C12 and Y112 in LigG-TD caused a complete loss of activity as well (Figure 3.27). In both possible Omega-class glutathione lyase mechanisms, Y113 is important for the binding of GSH-adduct, probably mediated by pi-pi interactions and a hydrogen bond between the tyrosine hydroxyl group and the keto group of the GSH-adduct in both mechanisms. In contrast, the role of C15 is more controversial. In the two-step mechanism postulated by Pereira et al.[71] for Omega-class glutathione lyases, C15 cleaves the thioether linkage of the GSH-adduct by nucleophilic attack, whereas in the one-step mechanism described here according to the mechanism postulated by Kontur et al.[60], C15 is seems to be responsible for the binding and activation of the second GSH molecule, which causes the thioether cleavage. Hence, independent of the catalytic mechanism, C15 is in both cases important for enzyme activity. In consequence, it is not surprising that C15 as well as Y113 are absolutely conserved in Omega-class glutathione lyases as confirmed by a multiple sequence alignment of the LigG PPR group members together with LigG-TD (Figure 4.6).

During protein engineering of LigG-TD also other residues with big influence on LigG-TD activity could be identified. Thus, amino acid exchange V11M was found to result in a 10% increase in LigG-TD activity in the conversion of racemic GS-VG (Table 3.10). V11 is located on a loop next to the essential amino acid C12. The mutation V11M might slightly shifts this loop, which would also move the residue C12 in direction to the thioether bond of the GSH-adduct (Figure 4.5).

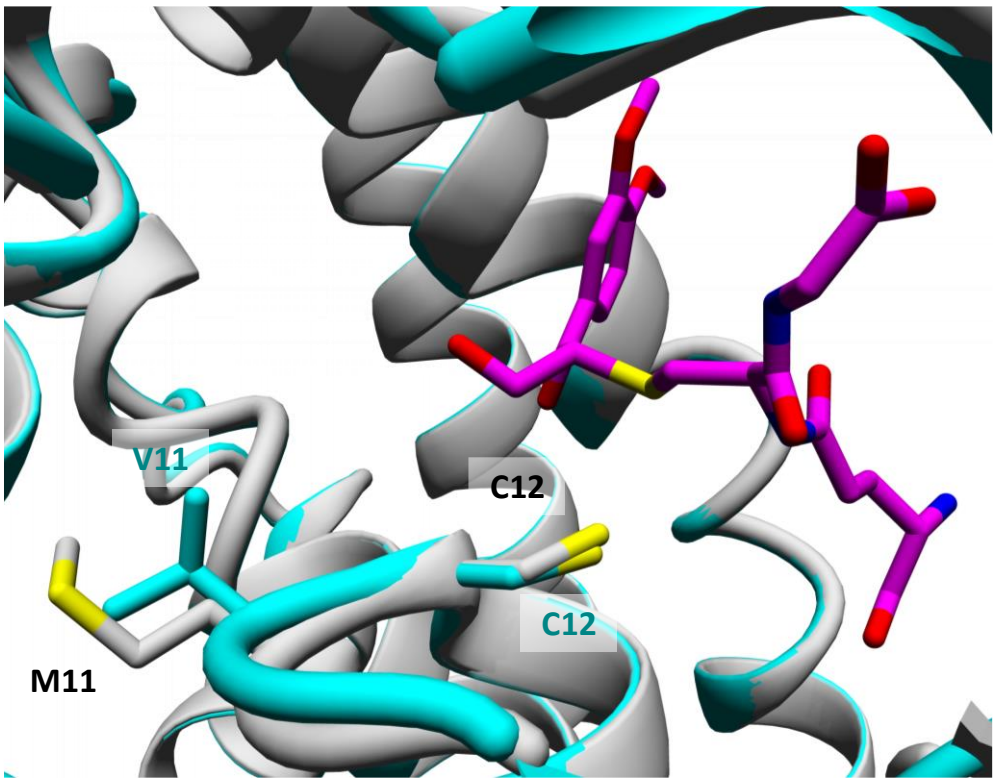


Figure 4.5 Possible influence of mutation V11M on the structure of LigG-TD. The structure of mutation LigG-TD V11M was energy minimized using Yasara with the force field Yasara 2. Shown is the superposition of the structures of LigG-TD wild-type (cyan) and mutant V11M (gray) with the amino acids M11, V11, C12, and substrate GS-(*R*)-VG (magenta) presented as sticks.

Also, mutations A55S and D104N (A56 and A 105 in LigG, respectively) resulted in an increased LigG-TD activity by 19 and 22%, respectively (Table 3.10). By the docking according to the one-step mechanism postulated by Kontur et al.[60], this can be explained by interactions of both amino acids with the GSH-adduct (Figure 3.16). In contrast, in the two-step mechanism postulated by Pereira et al.[71] no interactions between these two amino acids and the GSH-adduct are visible. However, D104 seems to be important for binding of the second GSH molecule in the second catalytic step, resulting in GSSG release, which could also explain the increased LigG-TD activity (Figure 4.3). Hence, to elucidate the influence and function of both residues with certainty, the mechanism of LigG-TD has to be confirmed first.

Another amino acid that needs to mentioned is N223 (also N223 in LigG). The amino acid exchange N223A nearly completely inactivated LigG-TD. Moreover screening of the SSM library exclusively at position N223 for active clones revealed wild-type hits (Figure 3.27 and Table 3.8), which indicates that this amino acid will be essential as well. This is supported by the multiple sequence alignment in Figure 4.6, which shows that N223 is conserved in all Omega-class glutathione lyases. The possible functional role of this amino acid can also be

deduced from substrate dockings (Figure 3.16). In the docking for both possible mechanisms, N223 interacts with the GSH-adduct. In case of the one-step mechanism according to the mechanism postulated by Kontur et al.[60], N223 additionally interacts with the GSH molecule as well. Furthermore, in the two-step mechanism postulated by Pereira et al.[71] N223 is also involved in the binding of the second GSH molecule in the second catalysis step. Therefore, also N223 seems to play a key role for LigG-TD catalysis.

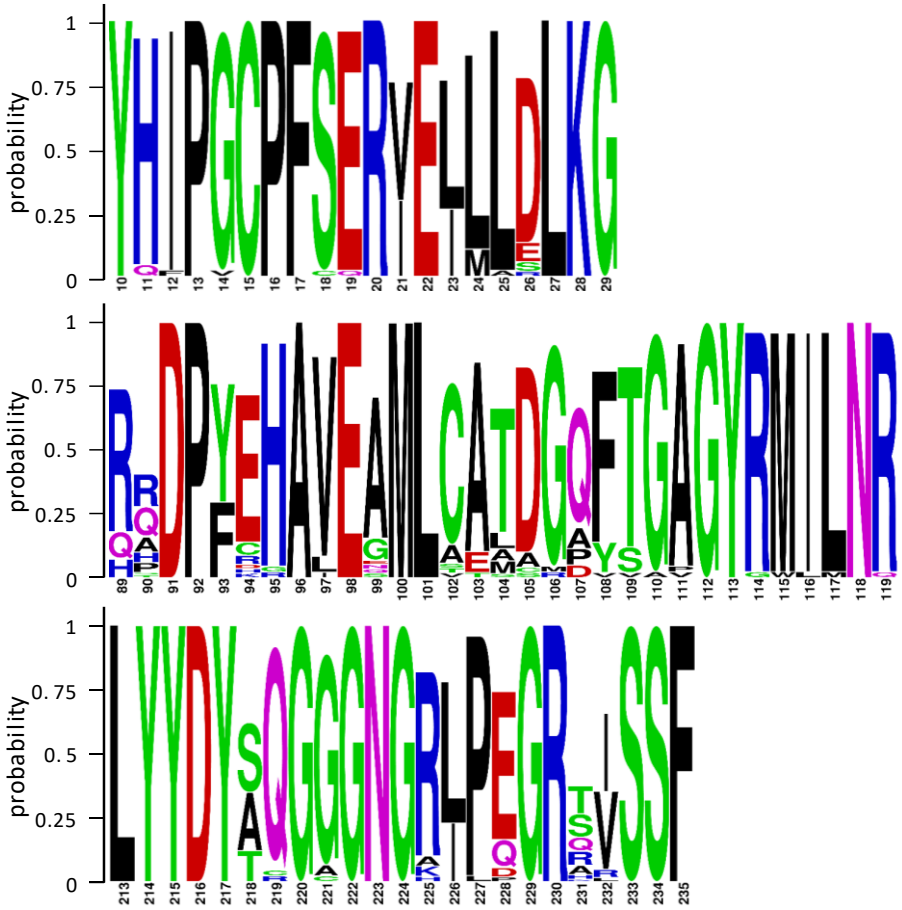


Figure 4.6 WebLogos (section 2.3.3.3) of sections from the multiple sequence alignment of Omega-class glutathione lyases, aligned with webPRANK. All sequences grouped together with LigG by the PPR algorithm (GenBank accession numbers of all sequences in the PPR group are listed in section 6.4) as well as LigG-TD were aligned. The numbering is according to the LigG sequence.

In general, more research is are required to better understand which amino acids are involved in the catalytic mechanism of LigG-TD and what is their specific function. For this, conformation of the catalytic mechanism will be the first step. Additionally, in silico reaction modeling using QM/MM models would be helpful to gain additional information about individual active site residues. This would also help to focus further protein engineering projects on amino acid residues which are not essential for catalysis but could positively influence enzyme activity and/or selectivity.

5 Conclusion and future perspectives

Lignin has a huge potential to become the major renewable resources for aromatic compounds in the future. A selective depolymerization of lignin will be the key step for the utilization of this potential. The glutathione-dependent lignin-degrading β-etherases and glutathione lyases, which were the topic of this thesis, demonstrated already high efficiency in the valorization of lignin.[32,34,63]

With classical database mining approaches and the peptide pattern recognition algorithm, 96 new putative β-etherases were identified and a representative set of 13 enzymes was further characterized to confirm their β-etherases activity. Several of the new β-etherases are even more active than the previously known enzymes. Hence, the use of these enzymes could improve current lignin degradation processes, as they are performed with less active β-etherases so far.

Based on multiple sequence alignments and the peptide pattern recognition analysis, a high number of conserved amino acids were identified among β-etherases. Using structural and functional analyses, several of these conserved amino acids could be correlated to a possible catalytic or structural function. This can be used in further investigations, experimentally or for in silico calculations, to better understand the catalytic mechanism of β-etherases and the involved amino acids. Also, in the case of the catalytic mechanism of glutathione lyases, this thesis delivers new aspects to investigate the catalytic mechanism of Omega-class glutathione lyases further and to determine the catalytic mechanism of LigG-TD in particular.

Phylogenetic analyses of β-etherases and glutathione lyases revealed new homologs, which can be interesting for further biochemical characterization. Especially in the case of hetero-dimeric β-etherases and Nu-class glutathione lyases, these homologs are of interest, as other approaches such as peptide pattern recognition were not successful to identify those enzymes. Moreover, homologs of Nu-class glutathione lyases are very interesting, since these enzymes seem to be the best suited glutathione lyases known for lignin valorisation based on their high activity and low stereoselectivity. Therefore, characterization of further members will increase the knowledge about this group of enzymes and could lead to the identification of even better enzymes.

Additionally, a phylogenetic cluster of LigF-type, NaLigF-2-type, and hetero-dimeric β-etherases, containing many other uncharacterized enzymes, was identified in the phylogenetic tree of β-etherases. Since this phylogenetic branch carries three types of lignin-

degrading enzymes, other, so far unknown types of lignin-degrading β-etherases might be found among the uncharacterized enzymes with possible new catalytic characteristics.

Glutathione-dependent lignin-degrading enzymes are also applied in the degradation of lignin polymer for the production small aromatic compounds in laboratory scale. For such applications, enzymes often need to be optimized by protein engineering, especially for industrial processes. In contrast to β-etherases, in the case of glutathione lyases no activity assay for high-throughput screening of mutant libraries was known. Therefore, an indirect activity assay for glutathione lyases was developed which enabled the identification of LigG-TD mutants with improved activity. The best LigG-TD variant is around 20% more active than the wild-type after in one round of protein engineering. This glutathione reductase based high-throughput screening assay can be important in the future to optimize also other glutathione lyases, e.g. Nu-class enzymes, for applications in lignin depolymerization.

Overall, this thesis has led to new insights into glutathione-dependent lignin-degrading enzymes and the identification of novel or optimized enzymes. Apart from that, the following continuative tasks need to be addressed in future work.

- The production of β-etherases and glutathione lyases, including expression and purification, has to be transferred to larger scale to produce sufficient amounts of biocatalysts for possible industrial applications. Especially the expression of the enzymes using bioreactors would be important. Here, the optimized protocols for shake flask expressions and small-scale purification, that have been established in this thesis, are a good basis.
- The high number of identified putative β-etherases and glutathione lyases are a great starting point for further studies and the identification of additional excellent biocatalysts. The novel enzymes could be more effective in lignin degradation than currently known ones. Additionally, the performed phylogenetic analyses of β-etherases and glutathione lyases might also lead to the identification of new groups of GSH-dependent lignin degrading enzymes.
- The established glutathione reductase based high-throughput assay could be further optimized, e.g. by using a robotic platform, and then be used for the optimization of glutathione lyases. Here, members of the Nu-class seem to be the best suited glutathione lyases for possible applications in lignin depolymerization and are a good target for further improvements.
- Important for the protein engineering of β-etherases and glutathione lyases is also a detailed understanding of their catalytic mechanisms to make the engineering more effective. Regarding this, more experimental investigations as well as in silico reaction modeling will be required in the future.

- Glutathione-dependent lignin-degrading enzymes in general are not used in industrial processes until now. Likewise, also lignin polymer is not utilized in industry as a renewable starting material so far. Therefore, a very big task for the future will be to develop commercially competitive processes either for lignin depolymerization, e.g. for the production of monomers for biobased plastics, or for the conversion of industrially relevant small molecules by β-etherases and glutathione lyases. The focus should lie in both cases on the production of high value compounds since biocatalytic processes are usually quite costly and, in the case of bulk chemicals, can hardly compete commercially with petrol-based chemical processes.

6 Appendix

6.1 DNA and Protein sequences of expressed proteins

All genes were cloned in pET-28a with the restriction enzymes NdeI, HindIII. The cutting sites are coloured in red whereas the his-tag is highlighted in blue.

pET28 LigE NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCACGCAATAATACCATCACCCTGTAT
GATCTGCAGCTGGAAAGCGGTTGTACCATTAGCCCGTATGTTTGGCGTACCAAATATGCCCTGAAACACAAAGGCTTCGATATTGATATT
GTTCCGGGTGGTTTTACCGGTATTCTGGAACGTACCGGTGGTCGTAGCGAACGTGTTCCGGTTATTGTTGATGATGGTGAATGGGTTCTG
GATAGCTGGGTTATTGCAGAATATCTGGATGAGAAATATCCGGATCGTCCGATGCTGTTTGAAGGTCCGACCCAGAAAAATCTGATGAA
ATTCCTGGATAATTGGCTGTGGTCAACCGCAGTTGGTCCGTGGTTTCGTTGTTATATCCTGGATTATCATGATCTGAGCCTGCCGCAGGAT
CGTGATTATGTTCGTTGGAGCCGTGAACAGTGGTTTCTGGGAGGTCAGCGTCTGGAAGATGTTCAGGCAGGTCGTGAAGATCGTCTGCC
GCTGGTTCCGCCTACCCTGGAACCGTTTCGTCGCATTCTGGCCGAAACCAAATGGCTGGGTGGTGATCAGCCGAATTTTGCAGATTATAG
CGCACTGGCAGTTTTTCTGTGGACCGCAAGCGTTGCACGTACCCCTCCGCTGACCGAAGATGATCCGCTGCGTGATTGGCTGGATCGTGG
TTTTGACCTGTTTGATGGCCTGGGTCGTCATCCGGGTATGAATCCGCTGTTTGGTCTGAAACTGCGTGAAGGTGATCCGGAACCTTTTGTT
CGTCAGACCGGTCCGGCTGGTGCGGGTGGTCAGGCACTGAATAAAGGTCCGCAGACCACCAAAATGCCTCCGCGTGTTGCAGAAAAAG
CAGATTAAAAGCTT

MGSSHHHHHHSSGLVPRGSHMARNNTITLYDLQLESGCTISPYVWRTKYALKHKGFDIDIVPGGFTGILERTGGRSERVPVIVDDGEWVLDSW
VIAEYLDEKYPDRPMLFEGPTQKNLMKFLDNWLWSTAVGPWFRCYILDYHDLSLPQDRDYVRWSREQWFLGGQRLEDVQAGREDRLPLVPP
TLEPFRRILAETKWLGGDQPNFADYSALAVFLWTASVARTPPLTEDDPLRDWLDRGFDLFDGLGRHPGMNPLFGLKLREGDPEPFVRQTGPA
GAGGQALNKGPQTTKMPPRVAEKAD

pET28 LigE-NA NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCAGCCAATAACACCATCACCTTTTAT
GATCTGGCACTGAGTACCGGTGCAACCATTAGCCCGTTTGTTTGGGCAACCAAATATGCCCTGAAACACAAAGGTTTTGATCTGGATGTT
GTTCCGGGTGGTTTTACCGGTATTCTGGAACGCACCGGTGGCAAAACCGAACGTCTGCCTGCAATTGTTGATGATGGTGAATTTGTTCTG
GATAGCTGGGGTATTGTTGAATATCTGGACGCAAAATATCCGGATCGTCCGGTTCTGATTCCGCATGAAAGCGTTGCAGCAACCCTGAAA
GCACTGGATAATTGGTTTTGGAATGCAGCAGTTGGTCCGTGGATGTTTTGTTTTTGTCAGGATTATCGTGATCTGAGCCTGCCGCAGGAT
CATGAATATGTTACCCATAGCCGTGAAAAAATGCTGGGTCGTAAACTGGAAGAGGTTCAGGCAGGTCGTGAAGAACGCCTGCCGAAAAT
TAGCGCAGCCCTGGAACCGCTGCGTGCAGCACTGGCACAGCATCAGTGGCTGGGTGGTAGCAGCCCGAATTATGCAGATTATCGCATTA
TGGGTGGCATTCTGTTTACCGCAAGCGTTTGTAAAACCCCTGTTCTGGCAAATGATGATCCGCTGCGCGATTGGATTGAACGTTGTCTGG
ATCTGTATGGTGGTCTGGGTCGCCATCCGGGTCTGTTTCCACTGTTTGGTCTGGAACAGCGTGAAGGTGATCCGGACCTGTTTAATCGTG
CAGCCGGTCAGGGTGGTATCTATAAACGTAATACGGGTCCGGAAAGCACCCGTGCAGAAACCCAGCGTATTACCGAAGGTATGAAAAA
GTAAAAGCTT

MGSSHHHHHHSSGLVPRGSHMAANNTITFYDLALSTGATISPFVWATKYALKHKGFDLDVVPGGFTGILERTGGKTERLPAIVDDGEFVLDSW
GIVEYLDAKYPDRPVLIPHESVAATLKALDNWFWNAAVGPWMFCFCQDYRDLSLPQDHEYVTHSREKMLGRKLEEVQAGREERLPKISAALE
PLRAALAQHQWLGGSSPNYADYRIMGGILFTASVCKTPVLANDDPLRDWIERCLDLYGGLGRHPGLFPLFGLEQREGDPDLFNRAAGQGGIYK
RNTGPESTRAETQRITEGMKK

pET28 LigE-NS NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCCAAAGATAATCGCATCACCCTGTA
TGATCTGCAGCTGGCAAGCGGTTGTACCATTAGCCCGTTTGTTTGGCGTACCAAATATGCACTGGCACATAAAGGTTTTGATGTGGATAT
TGTTCCGGGTGGTTTTACCGGTATTGCAGAACGTACCGGTGGTCGTAGCGAACGTGTTCCGGTTATTGTTGATGATGGTGAATGGGTTCT
GGATAGCTGGAAAATTGCCGAATATCTGGATGAACGTTATCCGGATCGTCCGATGCTGTTTGAAGGTCCGAGCATGAAACAGCTGACCA
AATTTCTGGATGCATGGCTGTGGCAGACCGCAATTGGTCCGTGGTTTCGTTGTTATATTCAGGATTATCATGATCTGAGCCTGCCTCAGG
ATCAGACCTATGTTCGTCATAGCCGTGAAACCATGTTTCTGGGTGGCAAAACCCTGGAAGAGGTTCAGGCAGGTCGTGAAGATCGTCTG
CCGCTGGTTCCGCCTACACTGGAACCGTTTCGTAAACTGCTGCGTGATACCCCGTGGCTGGGTGGTGATGCACCGAATTTTGCAGATTAT
ACCGCACTGGCCGTTTTTCTGTGGACCGCAAGCGTTGCAACCACCCCTCCGCTGACCGAAGATGATCCGCTGCGTGGTTGGCTGGATCGT

GGCTTTGATCTGTATGCAGGTCTGGGTCGTCATCCGGGTATGCATAGCCTGTTTGGTCTGCGTCTGCGTGAAGGTGATCCGGAACCTTTT
GCACGTGATGGTGCAGGTATTGAAGTTGCACCGGTTAATCGTGGCACCGCACGTAGCGCAGAAATGGCAGATTAAAAGCTT

MGSSHHHHHHSSGLVPRGSHMAKDNRITLYDLQLASGCTISPFVWRTKYALAHKGFDVDIVPGGFTGIAERTGGRSERVPVIVDDGEWVLDS
WKIAEYLDERYPDRPMLFEGPSMKQLTKFLDAWLWQTAIGPWFRCYIQDYHDLSLPQDQTYVRHSRETMFLGGKTLEEVQAGREDRLPLVPP
TLEPFRKLLRDTPWLGGDAPNFADYTALAVFLWTASVATTPPLTEDDPLRGWLDRGFDLYAGLGRHPGMHSLFGLRLREGDPEPFARDGAGIE
VAPVNRGTARSAEMAD

pET28 LigE179 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCAGCAAACAATACCGTGACCCTGTA
CGATCTGCAACTGGCATCCGGAGCAACAATCAGCCCGTTTGTGTGGGCAACGAAACTGGCGATTGCGCATAAAGGTTTGGATATGGAAA
TCGTTCCGGGTGGCTTTACCGGCATTGAGGAGCGTACGGGCGGCAAGACCCAGCGTTTGCCAGCGATCGTGGACGACGGCGAATGGAT
TCTGGATTCCTGGACCATTGCGGAATACCTGGATGAGAAATATCCAGACCGCCCTACCCTGATTGGTGACCCGGGTATCCGTCCGTCCGC
CCAAATGAACGAGGCCTGGTTGTGGCAGACCGCCGTTGGCCCGTGGATGACCTGTTATTTGGTCGCGTACCGCGATCGTAGCGTGCCGG
AAGACCACGAGTATGTTACCGCGACCCGTGAGACCATGTTCGGTGGTCGTAAATTGGAAGATATCATCGTCGGCCGTGAAGACCGTCTG
CCGCGTATTTCCGCGGACTTGGAGCTGATGCGCGGTGTTCTGCGCGAAAATAAGTGGTTTGGTGGTGATAGCCCGAATTACGCGGATCA
CCGTATGCTGGCGTGCTTCCTGTGGCTGGCGTCCGTGTGCGACACCCCGGCGCTGGCGGAAGATGATCCGCTGCGCGACTGGATTGACC
GTGGTTTCGATCTGTACGGTGGTATCGGTCGCCACCCGGGTCTGAGCAACATTTTCGGTCTGAAGCAGCGTGCGGGTGATCCGGACCTG
TTCAACCGCAACCCGATGGCGAACGGTCTGACCAGCCGTAATACCGGTGTTAAAAGCACCGCGGGTGAGACCGCGCGTATGAAGGGTC
AAAAGGCGGGTGCGTAAGCTT

MGSSHHHHHHSSGLVPRGSHMAANNTVTLYDLQLASGATISPFVWATKLAIAHKGLDMEIVPGGFTGIEERTGGKTQRLPAIVDDGEWILDS
WTIAEYLDEKYPDRPTLIGDPGIRPSAQMNEAWLWQTAVGPWMTCYLVAYRDRSVPEDHEYVTATRETMFGGRKLEDIIVGREDRLPRISAD
LELMRGVLRENKWFGGDSPNYADHRMLACFLWLASVCDTPALAEDDPLRDWIDRGFDLYGGIGRHPGLSNIFGLKQRAGDPDLFNRNPMA
NGLTSRNTGVKSTAGETARMKGQKAGA

pET28 LigE283 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCACGTAACAACACGATTACCTTCTAT
GATTTGGCGCTGTCCACCGGCGCGACCATCTCCCCGTTTGTTTGGGCAACGAAGTATGCATTGAAACACAAGGGTTTCGACCTGGACGTG
GTTCCGGGCGGCTTTACGAAAATCCCTGAACGCACGGGTGGCAAGACCGAGCGTTTGCCAGCTATTGTCGATGACGGCAAGTGGGTGTT
GGATTCCTGGGGTATCGTCGAGTACCTGGATGAAACCTATCCTGATCGTCCGATGCTGATTCCGCATCCGTCCGTGGCGGCCGTGACCCG
CGCCCTGGATGCGTGGTTCTGGGGTGCGGCGACCGGCCCGTGGATGCGCTGCTTTTGCGTGAACTACCGTGATCTGAGCAACCCGGAG
GATCATGAATACATCACCTATAGCCGTGAGAAAATGTTGGGTAAAACCCTGGAGGAAATGCAAGCTGGTTGGGAGGATCGCCTGCCAG
GTATCAGCGCTGCCCTGGAACCGTTGCGCATCGCCCTGCGTGAGGTTGACTACTTGGGTGGTGACGCGCCGAACTACGCGGATTACCGC
ATTCTGGGTAGCATCTTGTTCACCGCGAGCGTTGTAAAACCAGCCCGGTCCTGGCGGACGATGATCCGCTGCGTCCGTGGATTGACAATT
TGCTGGACATGTTCGGTGGTCTGGGTCGTCACCCGGGTCTGTTCCCGTTGTTTGGTCTGAAACAGCGTGAAGGTGACCCGCCGCTGTTCA
TGCCTGCGGGTATGGGTGGCATCCATAAGCGTAATACCGGTGTTGAGAGCACCCGTGCGGAGACCCGTCGTATTACCGAAGGTATGGC
GAAGAATTAAGCTT

MGSSHHHHHHSSGLVPRGSHMARNNTITFYDLALSTGATISPFVWATKYALKHKGFDLDVVPGGFTKIPERTGGKTERLPAIVDDGKWVLDS
WGIVEYLDETYPDRPMLIPHPSVAAVTRALDAWFWGAATGPWMRCFCVNYRDLSNPEDHEYITYSREKMLGKTLEEMQAGWEDRLPGISA
ALEPLRIALREVDYLGGDAPNYADYRILGSILFTASVCKTSPVLADDDPLRPWIDNLLDMFGGLGRHPGLFPLFGLKQREGDPPLFMPAGMGGI
HKRNTGVESTRAETRRITEGMAKN

pET28 LigE491 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGACCGATCGCCATAGCGTGACCCTGTA
CGATCTGAATCTGGCAAGCGGATGCACCATCTCCCCGTTTGTTTGGCGTACCAAGTATGCTTTGGCACACAAAGGTCTGGCTATTGATATT
GTTCCGGGCGGTTTCACCGGCATTCTGGAGCGTACCGGCGGTCGTAGCGAACGTTTGCCGGCGATCATCGACAATGGTGAATGGGTTTT
GGACAGCTGGCTGATTGCGGAGTACTTGGACGGCAAGTATCCAGAGCGCCCACTGTTCGAAGGCCCGTCCATGAAGGTGTTGACCAAGT
TCATTGACCAATGGTTGTGGCGTACCGCCATCGGTCCGTGGTTTCGTTGTTATATCCTGGATTACCATAATCTGAGCTACCCGCATGACCA
CGACTACATTCGTACCACCCGTGAAACCATGTTCTTGGGTGGCCAGAAACTGGAAGACGTGCAAGCAGGTCGTGAGGAACGTCTGCCTT
TGGTCCCTCCGCTGTTGGAGCCTTTGCGTCAACTGCTGCGTGACACCCCGTGGTTGGGTGGTGATACCCCGAACTACGCGGATTACTGCG
CCCTGGCGGTTTTCCTGTGGACCGCGAGCGTGGCGACCACGCCGCCGCTGACCGACGACGACCCGCTGCGTGATTGGCTGGATCGTGGT
TTTGACCTGTATGATGGTCTGGGTCGCCACCCGGGCATGCACAGCCTGTTTGGTCTGCAGCTGCGCGAAGGTGATCCGGCGCCTTTCCTG
AAAGCGGGTATCGGTACGGCGCCAGCGCCGGTTAATCGTGGTGCGGGTACCACGCCATTCCCGGAACCGGCGCGCGCTGCGTAAGCTT

MGSSHHHHHHSSGLVPRGSHMTDRHSVTLYDLNLASGCTISPFVWRTKYALAHKGLAIDIVPGGFTGILERTGGRSERLPAIIDNGEWVLDSW
LIAEYLDGKYPERPLFEGPSMKVLTKFIDQWLWRTAIGPWFRCYILDYHNLSYPHDHDYIRTTRETMFLGGQKLEDVQAGREERLPLVPPLLEPL
RQLLRDTPWLGGDTPNYADYCALAVFLWTASVATTPPLTDDDPLRDWLDRGFDLYDGLGRHPGMHSLFGLQLREGDPAPFLKAGIGTAPAP
VNRGAGTTPFPEPARAA

pET28 LigE760 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCCAAAGACAACAAAATTACGTTCTT
CGACCTGACGCACGAAAGCGGATGCACGACGAGCCCATTCGTGTGGGCTACGAAATACGCCGTCAAACATAAAGGTTTCGACTTGGACG
TGGTTCCGGGCGGTTTCACCGGCATTCTGGATCGTACGGGCGGCCGCTCCGAACGCCTGCCGGTTATCTGCGATGATGGCGAGTACGTT
CTGGATAGCTGGCTGATTGCAGAATATTTGGATGCCAAATACCCGGACCGTCCGACCTTGATCGGCGACCCGAGCGTGAAGGTGCTGAC
GCAATTCCTGGAGACGTGGTTGTGGAAGACCGTGGTCGGCCCTTGGGCGCGCTGTTTTGCGGTGCAATACCGTGACCGCTGTTTCCCGC
ACGACATTCAGTATATCACCGAGAGCCGTTTGCGCATGTGGGGTAAACCGATGGAAGAATTGATCGTGGGCCGCGAAGATGTGTTTCCT
AAAGTTTTGCCGGAATTGGAGCTGCTGCGTGGTATCCTGCGTGAGCACAAATGGCTGGGTGGCGAAACCCCGAACTATGCTGATTACCG
TGCGCTGGCTGTGTTCTTGTGGGCGGCGAGCGTTGCGGACACCCCGCCGATGACCGAGGACGATCCGTTGCGTGACTGGATCGATCGTG
GTTTCGACCTGTATGGTGGTCTGGGTCGTATTCCGGGTATGAGCCCGCTGTTCGGTCTGAAGATGCGTGAGGGTGACCCGGAGCCGTTC
GCGAAGGGTCCGTCCCTGGTGGCGGGTCTGGTTAAGCGTAATACCGGTCCGGCGTCCACCGCGGCGGAAACCGCGCACATTACGGGTA
AGGGTGAAAAGGCGCCGGCCTAAGCTT

MGSSHHHHHHSSGLVPRGSHMAKDNKITFFDLTHESGCTTSPFVWATKYAVKHKGFDLDVVPGGFTGILDRTGGRSERLPVICDDGEYVLDS
WLIAEYLDAKYPDRPTLIGDPSVKVLTQFLETWLWKTVVGPWARCFAVQYRDRCFPHDIQYITESRLRMWGKPMEELIVGREDVFPKVLPELE
LLRGILREHKWLGGETPNYADYRALAVFLWAASVADTPPMTEDDPLRDWIDRGFDLYGGLGRIPGMSPLFGLKMREGDPEPFAKGPSLVAGL
VKRNTGPASTAAETAHITGKGEKAPA

pET28 LigE889 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCACAGAACAACACCATCACCTTTTAT
GATCTGGCAATTTCCACCGGTGCCACGATCTCCCCGTTTGTGTGGGCTACGAAATACGCGTTGAAGCACAAGGGCTTTGACTTGGATGTC
GTGCCTGGTGGTTTCACGAAGATCCCAGAACGTACCGGTGGTGTCACGGAACGTTTGCCGGCTATTGTTGACGATGGCAAATGGGTGTT
GGACAGCTGGGGTATCGTTGAGTACTTGGATGCCACCTACCCTGATCGCCCGTTGTTGATTCCTCATCCATCCGTTGCGATTGTGACCCGT
GCTCTGGATGCGTGGTTCTGGCAGGTGGCGACCGGCCCGTGGATGCGTTGCAATTGCGTTTCCTATCGTGATTTGAGCAATCCGGAAGA
CCACGAGTACATCACCCATTCCCGCGAAAAGATGCTGGGTAAAACCCTGGAGGCGATGCAGGAGGGTTACGAAGACCGTTTGCCGGCG
ATTAGCGCGGCGCTGGAGCCGCTGCGTATCGCGCTGCGCGAGGGTGATTGGCTGGGTGGCGATAGCCCGAATTACGCGGACTATCGCA
TTCTGGGCAGCATTCTGTTTACCGCGAGCGTTTGTCAGAACAGCCCGGTTCTGGCGGAAGACGACCCGCTGCGCCCGTGGATCGAACGT
TGTCTGGATCTGTATGACGGTCTGGGTCGTCACCCGGGTCTGTTCCCGCTGTTCGGTCTGGAGCAACGCCCGCAAGACCCGCCGCTGTTC
ATGCCGCAAGGTCAAGGTGGTATTCACAAACGTAACACCGGTCCGGCGAGCACCAGCGCGGAGACCCAACGTATCACCGAAGGTATGG
CGAAAGCGTAAGCTT

MGSSHHHHHHSSGLVPRGSHMAQNNTITFYDLAISTGATISPFVWATKYALKHKGFDLDVVPGGFTKIPERTGGVTERLPAIVDDGKWVLDS
WGIVEYLDATYPDRPLLIPHPSVAIVTRALDAWFWQVATGPWMRCNCVSYRDLSNPEDHEYITHSREKMLGKTLEAMQEGYEDRLPAISAAL
EPLRIALREGDWLGGDSPNYADYRILGSILFTASVCQNSPVLAEDDPLRPWIERCLDLYDGLGRHPGLFPLFGLEQRPQDPPLFMPQGQGGIHK
RNTGPASTSAETQRITEGMAKA

pET28 LigE915 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCGCAAGACAATAAGATCACCTTCTA
CGACTTGGCAATTAGCACGGGTGCGACCATTAGCCCTTTCGTTTGGGCCACGAAGTACGCATTGAAACACAAAGGCTTTGATCTGGACGT
GGTGCCAGGCGGCTTTACCGGCATCCTGGAGCGCACGGGCGGCAAAACGGAACGCTTGCCAGCGATCGTGGATGATGGTACCTGGGTG
CTGGATTCCTGGGGCATTGTTGAATATCTGGATGCCACCTACCCGGACCGTCCGTTGCTGATTCCGCACGAGAGCGTCGCTGCCACGTTG
AAAGCGTTGGACCACTGGTTCTGGGGTGCTGCGGTCGGTCCGTGGATGCGCTGCTTTTGCGCGGATTACCGTGACTTGAGCCTGCCTCA
GGATCACGAATATATCACCCATTCCCGTGAAAAGATGTTGGGTCGTAAGTTGGAGGAGATTCAGGCGGGTCGTGAGGACCGTTTGGCTG
GTATTAGCGCGGCGTTGGAACCGCTGCGTAGCGCGCTGGGTCAACATGCGTGGCTGGGTGGTAGCAGCCCGAATTACGCGGATTATCG
CATTCTGGGCGGTTTTCTGTTCACCGCGTCCGTGTGTAAAACCCCGGTCCTGGCGAGCGATGATCCGCTGCGTGACTGGCTGGACCGTTG
TCTGGATCTGTATGGTGGTCTGGGTCGCCATCCGGGTCTGTTCCCGCTGTTCGGTCTGGTGGAACGCGAAGGCGATGTTCCGGCGTTTAC
CCGTCAGGGTGGTTTGGGTGGTATCTACAAACGCAACACCGGTCCGGCGAGCACCCAAGCGGAAACCCAGAAAATCACCCAAGGTATG
AAGCAACCGGCGTAAGCTT

MGSSHHHHHHSSGLVPRGSHMAQDNKITFYDLAISTGATISPFVWATKYALKHKGFDLDVVPGGFTGILERTGGKTERLPAIVDDGTWVLDS
WGIVEYLDATYPDRPLLIPHESVAATLKALDHWFWGAAVGPWMRCFCADYRDLSLPQDHEYITHSREKMLGRKLEEIQAGREDRLAGISAALE

PLRSALGQHAWLGGSSPNYADYRILGGFLFTASVCKTPVLASDDPLRDWLDRCLDLYGGLGRHPGLFPLFGLVEREGDVPAFTRQGGLGGIYKR
NTGPASTQAETQKITQGMKQPA

pET28 LigP NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCCAAAGACAACAAGATCACCATTTA
TGATCTGGCACTGGCAAGCGGTGCAACCATTAGCCCGTTTGTTTGGGCAACCAAATATGCAATTGCCCACAAAGGTTTTGAGCTGGATAT
TGTTCCGGGTGGTTTTAGCGGTATTCCGGAACGTACCGGTGGTGTTACCGAACGTCTGCCTGCAATTGTTGATGATGGTAAATGGGTTCT
GGATAGCTGGCTGATTGCAGAATATCTGGATGAAACCTATCCTGAACGTCCGACCCTGATTCCGCATGCAAGCGTTAAAGCACTGACCCA
GGGTATGGAAGCATGGCTGTGGGGTGCAGCAATTAGTCCGTGGATGACCTGTTTTATCAAACAGTATCGTGATCGTAGCCTGCCGCAGG
ATCATGAATATGTTACCACCAGTCGTGAACGCATGTTTGGTCGTAAAATTGAAGATATTATTGTGGGTCGCGAAGATCGCATTCCGAAAG
TTCCGCCTACCCTGCAGCTGCTGCGTAATGTTCTGGCAGAAAACAAATGGCTGGGTGGTGATACCCCGAATTATGCAGATTTTCGTCTGC
TGGCAGTTTTTCTGTTTACCGCCAGCGTTGCAGATACACCGGTTCTGACCGATGATGATCCGCTGCGTGATTGGATTGAACGTGGTTTTG
ATCTGTATGGTGGTCTGGGTCGTCATCCGGGTCTGAGCCCGATTTTTGGTCTGCAACTGCGTGAAGGTGATCCGGAACCGTTCATTAAAG
GTGGTGCAGTTGGTGGCCTGGCAACCCGTAATACCGGTCCGAAAAGCACCGCAGCAGAAACCGCACGTCTGAAAGGTGAAAAAGCTCC
GGCAGCATAAAAGCTT

MGSSHHHHHHSSGLVPRGSHMAKDNKITIYDLALASGATISPFVWATKYAIAHKGFELDIVPGGFSGIPERTGGVTERLPAIVDDGKWVLDSW
LIAEYLDETYPERPTLIPHASVKALTQGMEAWLWGAAISPWMTCFIKQYRDRSLPQDHEYVTTSRERMFGRKIEDIIVGREDRIPKVPPTLQLLR
NVLAENKWLGGDTPNYADFRLLAVFLFTASVADTPVLTDDDPLRDWIERGFDLYGGLGRHPGLSPIFGLQLREGDPEPFIKGGAVGGLATRNT
GPKSTAAETARLKGEKAPAA

pET28 LigF NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGACCCTGAAACTGTATAGCTTTGGTCC
GGGTGCAAATAGCCTGAAACCGCTGGCAACCCTGTATGAAAAAGGTCTGGAATTTGAACAGGTTTTTGTGGATCCGAGCAAATTTGAGC
AGCATAGCGATTGGTTCAAAAAGATTAATCCGCGTGGTCAGGTTCCGGCACTGTGGCATGATGGTAAAGTTGTTACCGAAAGCACCGTG
ATTTGTGAGTATCTGGAAGATGTTTTTCCGGAATCAGGTAATAGTCTGCGTCCGGCAGATCCTTTTAAACGTGCAGAAATGCGTGTTTGG
ACCAAATGGGTTGATGAATATTTCTGTTGGTGCGTTAGCACCATTGGTTGGGCATTTGGTATTAAAGCAATTGCCCAGAAAATGAGCGAC
GAAGAATTTGAGGAACACATCAATAAAAACGTTCCGATCCCGGAACAGCAGCTGAAATGGCGTCGTGCACGTAATGGTTTTCCGCAAGA
AATGCTGGATGAAGAATTCCGTAAAGTGGGTGTTAGCGTTGCACGCCTGGAAGAAACCCTGAGCAAACAGGATTATCTGGTTGATACCG
GTTATAGCCTGGCCGATATTTGTAATTTTGCCATTGCAAATGGTCTGCAGCGTCCTGGTGGTTTTTTTGGTGATTATGTGAATCAAGAAAA
AACCCCTGGTCTGTGTGCATGGCTGGATCGTATTAATGCACGTCCTGCAATCAAAGAGATGTTTGAAAAAAGCAAACGCGAGGACCTGC
TGAAACGTCAGAACGAAAAAGTTGCATAAAAGCTT

MGSSHHHHHHSSGLVPRGSHMTLKLYSFGPGANSLKPLATLYEKGLEFEQVFVDPSKFEQHSDWFKKINPRGQVPALWHDGKVVTESTVICE
YLEDVFPESGNSLRPADPFKRAEMRVWTKWVDEYFCWCVSTIGWAFGIKAIAQKMSDEEFEEHINKNVPIPEQQLKWRRARNGFPQEMLDE
EFRKVGVSVARLEETLSKQDYLVDTGYSLADICNFAIANGLQRPGGFFGDYVNQEKTPGLCAWLDRINARPAIKEMFEKSKREDLLKRQNEKVA

pET28 LigF-NA NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTGAAACTGTATAGCTTTGGTCCGGC
AGCAAATAGCATGAAACCGCTGCTGACCGTTTTTGAAAAAGGCCTGGATGTTGAAAAACATCGTCTGGATCCGGCAAAATTTGAACATCA
TACCGATTGGTTCAAAGCCATTAATCCGCGTGGTCAGGTTCCGGCACTGGTTGATGGTGATAAAGTTGTTACCGAAAGCACCGTGATTTG
CGAATATCTGGAAGATGAATATCCGACCGAAGTTGCACTGCGTCCTGCAGATAGTTTTGGTAAAGCACAGATGCGTATTTGGACAAAAT
GGGTGGATGAGTATTTTTGTTGGTGCGTTAGCACCATTGGTTGGCATCGTTATGTTGGTAATATGGTGAAAAGCCTGAGTGATGCCGAAT
TTGAAGAAAAAGTTAAAGCCATCCCGGTGATTGAACAGCAGGTTAAATGGCGTCGTGCACGTGAAGGTTTTCCGCAGGATATGCTGGAT
GAAGAAATGCGTAAAATTGCCTATAGCGTGCGTAAACTGGATGATCACCTGGCAGATCATGAATGGCTGGTTCCGGGTCAGTATACCCT
GGCCGATATTTGTAATTTTGCAATTGCCAATGGTATGCAGTTTGGTTTTGCAGAACTGGTGAATAAACAGGATACACCGCATCTGGTTCG
TTGGATTGAGCAGATTAATGAACGTCCTGCCGTGAAACAAATGTTTGCACAGGTTGAACTGGAAAAACTGGGTCCGCGTGAATAAAAGC
TT

MGSSHHHHHHSSGLVPRGSHMLKLYSFGPAANSMKPLLTVFEKGLDVEKHRLDPAKFEHHTDWFKAINPRGQVPALVDGDKVVTESTVICEY
LEDEYPTEVALRPADSFGKAQMRIWTKWVDEYFCWCVSTIGWHRYVGNMVKSLSDAEFEEKVKAIPVIEQQVKWRRAREGFPQDMLDEEM
RKIAYSVRKLDDHLADHEWLVPGQYTLADICNFAIANGMQFGFAELVNKQDTPHLVRWIEQINERPAVKQMFAQVELEKLGPRE

pET28 LigF-NS NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTGACCCTGTATAGCTTTGGTCCGGG
TGCAAATAGCCTGAAACCGCTGCTGGCACTGTATGAAAAAGGTCTGGAATTTACACCGCGTTTTGTTGATCCGACCCGTTTTGAACATCAT

GAAGAATGGTTCAAAAAGATCAATCCGCGTGGTCAGGTTCCGGCACTGGATCATGATGGTCATATCATTACCGAAAGCACCGTGATTTGT
GAGTATCTGGAAGATGCATTTCCGGAAGCTCCGCGTCTGCGTCCGGTGGATCCGGTTATGATTGCAGAAATGCGTGTTTGGACCAAATG
GGTGGATGAATATTTCTGTTGGTGTGTTAGCACCATTGGTTGGGAACGTATGATTGGTCCGATGGCACGTGCACTGAGCGACGAAGAAT
TTGAAGCAAAAGTTGCACGTATTCCGGTTCCAGAACAGCGTACAAAATGGCGTACCGCACGTACCGGTTTTCCGAAAGAAGTTCTGGAC
GAAGAGATGCGTAAAATTGGTGTTAGCGTTAATCGTCTGGAAACCCGTCTGGCAGAAAGCCCGTGGCTGGCAGGCGAAAACTTTAGCCT
GGCAGATGTTTGTAATTTTGCCATTGCAAATGGTATGCAGAACGGCTTTAGCGATATTGTGAATCGTGAAGCAACACCGCATCTGGTTGC
ATGGATTGAAAAAATCAATGATCGTCCGGCATGCAAAGCAATGTTTGCCAATAGCAAAAGCGAATTTGCAGATCGCGGTCAGAAAGTTA
CCGCATAAAAGCTT

MGSSHHHHHHSSGLVPRGSHMLTLYSFGPGANSLKPLLALYEKGLEFTPRFVDPTRFEHHEEWFKKINPRGQVPALDHDGHIITESTVICEYLED
AFPEAPRLRPVDPVMIAEMRVWTKWVDEYFCWCVSTIGWERMIGPMARALSDEEFEAKVARIPVPEQRTKWRTARTGFPKEVLDEEMRKIG
VSVNRLETRLAESPWLAGENFSLADVCNFAIANGMQNGFSDIVNREATPHLVAWIEKINDRPACKAMFANSKSEFADRGQKVTA

pET28 LigF008 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGTTGAAGCTGTATTCCTTCGGCCCTGGC
GCGAACAGCCTGAAGCCGATGTTGACGTTGTTTGAGAAGGGCCTGGAGTACGAGCAACATCAGTTGAACCCGGCAAAATTCGAGCATC
ATAGCGACTGGTACAAGGCAGTGAACCCGCGTGGTCAAGTCCCAGCTTTGGACGACAATGGTCGTATTGTGACGGAATCCACCGTTATC
TGTGAATACTTGGAAGATGCCCACCCGACCGCGGTTAAACTGCGCCCGGATGACCCGTACGACCGTGCTCAAATGCGTGTGTGGACCAA
ATGGGTGGATGAGTATTTCTGCTGGTGCGTGTCCACGATCGGCTGGCACCGTGGTGTCCGTTTCATGGCCCAACAATTGAGCGACGCGG
AGTTCGAGGAACACCTGAAGAAAATCCCTATCCCGGAACAGCAGGTCAAATGGCGCCGTGCGCGCGAAGGTTTTCCTCAGGATCTGCTG
GATGAAGAGATGCGTAAAATTGGTGTGAGCGTGCGCAAACTGGACGACCACCTGGCGGATAACGAATGGCTGGCGGGTGGTATGTTCA
GCCTGGCGGATATTTGTAATTTCGCAATCGCGAATGGTATGGAGGTTGGTTTCGCGGACCAGGTTAACAAGCAAGATACCCCACACCTG
GTTCGTTGGATTGAGCAAATTAATGCGCGTCCGAAGGTTCAGGAAATGTTTGCGGCGGTTCCGCGTGAACGTCTGGGTCCGCCTAAGTA
AGCTT

MGSSHHHHHHSSGLVPRGSHMLKLYSFGPGANSLKPMLTLFEKGLEYEQHQLNPAKFEHHSDWYKAVNPRGQVPALDDNGRIVTESTVICEY
LEDAHPTAVKLRPDDPYDRAQMRVWTKWVDEYFCWCVSTIGWHRGVRFMAQQLSDAEFEEHLKKIPIPEQQVKWRRAREGFPQDLLDEE
MRKIGVSVRKLDDHLADNEWLAGGMFSLADICNFAIANGMEVGFADQVNKQDTPHLVRWIEQINARPKVQEMFAAVPRERLGPPK

pET28 LigF215 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGTTGACCCTGTACAGCTTCGGCCCATCC
GCAAATAGCCTGAAACCGCTGTTGACGCTGTACGAGAAGGGTCTGGATTTCGAAAGCCGTTACCTGAATCCTGCACGCTTCGAACAACAT
GAAGATTGGTTTAAGCAGATCAACCCGAATGCTCAGGCACCTGCGCTGGATCACGATGGTCGTATCATTACCGAGAGCACCGTTATCTGT
GAGTACCTGGAGGATTTGTTTCCGGGTACCCCGCGTTTGCGTCCGGAAGATCCGTATCAGCGCGCTCAGATGCGCATTTGGACGAAATG
GGTTGACGAATACTTCTGCTGGTGTGTGAGCACCATCGGCTGGGAACGTCGCATCAGCGGCATTGCGCGTTCCGTTTCCGAAGAAGAAT
TCGAGCGTCTGGTTGCGCGTATCCCGCTGAAAGAGCAACAATTGAAGTGGCGTAACGCGCGTAATGGTTTCGACAAGAAAGTGTTGGAA
GAGGAAATGCGTAAAATTGACTATAGCGTGAAGAAACTGGAGGCTCGTTTGTCCCAATCCCCGTGGCTGGCTGGTGACTCCTACACGCT
GGCGGACATCTGCAACTTCGCGATTGCGCACACGATGGATAGCGGTTTTCCGGAGCTGATTAACCCGCAGGCGACCCCACATATTTTGG
ATTGGTTGGCGCGTATTCGCGAGCGTCCGGCGTGCCGTACCATGTATGCTAACGCGCCTGTTCGCTAAGCTT

MGSSHHHHHHSSGLVPRGSHMLTLYSFGPSANSLKPLLTLYEKGLDFESRYLNPARFEQHEDWFKQINPNAQAPALDHDGRIITESTVICEYLE
DLFPGTPRLRPEDPYQRAQMRIWTKWVDEYFCWCVSTIGWERRISGIARSVSEEEFERLVARIPLKEQQLKWRNARNGFDKKVLEEEMRKIDY
SVKKLEARLSQSPWLAGDSYTLADICNFAIAHTMDSGFPELINPQATPHILDWLARIRERPACRTMYANAPVR

pET28 LigF729 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTGAAACTGTATTCCTTTGGCCCGGCC
GCCAACTCCATGAAGCCTTTGCTGACCCTGTATGAGAAGGGCACCCCGTTCGAAGGTCACCGCTTGAACCCGGCTATCTTTGAACATCAT
CAGGACTGGTTTAAAGCGATCAATCCGCGTGGCCAAGTCCCGGCGCTGGTCGACGGTGATAAAGTGATCACGGAAAGCACGGTTATTTG
CGAATATCTGGAGGACGAACATCCGGGTGAGGTTAAATTGCGCCCGGCTGATTCCTACGGCCGCGCACAAATGCGTATCTGGACCAAGT
GGGTGGATGAATACTTTTGTTGGTGTGTGAGCACCATTGGTTGGCACCGTTACGTCGGTAATATGGTGAAGAGCTTGAGCGACGAGGCG
TTTGAGGAAAAGGTTAAGGCGATTCCGGTTTTCGAACAACAAGTGAAATGGCGTCGTGCACGCGAGGGTTTCCCGCAGGATCTGCTGGA
TGAAGAAATGCGCAAAATTGCGTTCAGCGTGACCCGTCTGAATGACCACTTGGCGGATCACGAGTGGCTGGCGGGTGATATGTTCACCC
TGGCGGACATTTGCAATTTCGCGATTGCGAACGGTATGAACTACAGCTTCCCTGAGCTGGTTAACGAGGCGGATGCGCCGCACCTGGTG
CGTTGGATCGCGCAAATCAACGCGCGTCCGGCGGTTCAGAAGATGTTCGCGGAAGTTCCAATGGAGAAACTGCGCCCGCAGGACTAAG
CTT

MGSSHHHHHHSSGLVPRGSHMLKLYSFGPAANSMKPLLTLYEKGTPFEGHRLNPAIFEHHQDWFKAINPRGQVPALVDGDKVITESTVICEYL
EDEHPGEVKLRPADSYGRAQMRIWTKWVDEYFCWCVSTIGWHRYVGNMVKSLSDEAFEEKVKAIPVFEQQVKWRRAREGFPQDLLDEEMR
KIAFSVTRLNDHLADHEWLAGDMFTLADICNFAIANGMNYSFPELVNEADAPHLVRWIAQINARPAVQKMFAEVPMEKLRPQD

pET28 LigF755 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGTTGACCTTGTATAGCTTCGGCCCAGCC
GCAAACAGCCTGAAACCATTGTTGGCCTTGTATGAAAAGGGCTTGGACTTCACCCCGCGCTTTGTTGATCCTACCCGTTTCGAACACCACG
AGGACTGGTTTAAAGCCATCAACCCGCGCGGTCAGGTCCCGGCGTTGGATCACGATGGTCATATTATCACGGAGTCCACGGTGATTTGT
GAATATTTGGAGGATGCCTTTCCGGACGCCCCACGCCTGCGTCCGACCGACCCTGTGCAAATCGCAGAAATGCGTGTTTGGACCAAATG
GGTTGACGAATACTTCTGCTGGTGTGTGTCCACCATCGGTTGGGAGCGCATGATTGGTCCGATGGCACGCAAGTTGAGCGACGAGGAAT
TCGAGGAGAAGTTGAAACATATCCCGATTCCGGAACAACAAGCGAAATGGCGCAGCGCGCGTGCGGGTTTCCCGCAGGCGGTGCTGGA
CGAGGAAATGCGTAAGATTCGTGTTTCCATCGAGCGTCTGGAAAAGCGTCTGAGCCAGAGCCCGTGGCTGGCTGGTGAAGTGTACACCC
TGGCGGATATTTGTAATTTTGCGATTGCGAACGGTATGCAATTCGGTTACGCGGATATTGTCAATCCGGATGCGACCCCGAACCTGGTTG
CGTGGATCGATCGTATCAACGCGCGTCCGGCAGCGCAGGCGATGTTTGCGAATAGCCGTAGCGAGATGCCGCCTCGTCCGGCGGCGGC
GACCGCTGCGTAAGCTT

MGSSHHHHHHSSGLVPRGSHMLTLYSFGPAANSLKPLLALYEKGLDFTPRFVDPTRFEHHEDWFKAINPRGQVPALDHDGHIITESTVICEYLE
DAFPDAPRLRPTDPVQIAEMRVWTKWVDEYFCWCVSTIGWERMIGPMARKLSDEEFEEKLKHIPIPEQQAKWRSARAGFPQAVLDEEMRKI
RVSIERLEKRLSQSPWLAGEVYTLADICNFAIANGMQFGYADIVNPDATPNLVAWIDRINARPAAQAMFANSRSEMPPRPAAATAA

pET28 LigF921 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGTTGAAACTGTACAGCTTCGGCCCAGG
CGCAAATTCCTTGAAACCTATGCTTACCCTGCATGAGAAGGGCTTGGATTACGAACAGGTGCTGCTGGACCCGCGTAAGTTTGAACATCA
CAGCGATTGGTTCAAGAAGATTAACCCGCGCGGTCAGGTTCCGGCACTGGAGGATCGTGGTCATATTATCACCGAATCCACGGTGATCT
GTGAATATCTGGAAGACGAGTGGCCGACGGACGTCAAGTTGCGCCCGGACACCAGCTGGGAGCGCGCTGACATGCGCGTGTGGACCAA
ATGGGTGGATGAGTACTTTTGCTGGTGCGTTTCCACCATCGGTTGGCACCGCTATGTGGGTAACATGGTCAAAGGTTTGAGCGATGAAG
AGTTTGAAGAAAAGGTTAAGAACATCCCGGTGGTTGAACAGCAGGTTAAGTGGCGTCGTGCTCGTGAGGGTTTCCCGCAGGAGATGTT
GGATGAGGAGATGCGTAAAATTGCGTACTCCGTCCGTAAGTTGGACGACCACTTGCGCGACAACGAGTGGTTGGTTCCGGGTATGTATA
CCCTGGCGGACATCTGTAATTTTGCGATCGCGAATGGTATGCAACACGGTTATCCGGAACTGGTTAATAAAGAAGACACCCCGGGTCTG
CTGCGTTGGATTGAGCAAATTAATGAGCGTCCGGCGGCGCAAAAGATGTTCGCGGACGTGCCGCGTGAGATTCGTCACGAGGATGCGA
AGAAATAAGCTT

MGSSHHHHHHSSGLVPRGSHMLKLYSFGPGANSLKPMLTLHEKGLDYEQVLLDPRKFEHHSDWFKKINPRGQVPALEDRGHIITESTVICEYLE
DEWPTDVKLRPDTSWERADMRVWTKWVDEYFCWCVSTIGWHRYVGNMVKGLSDEEFEEKVKNIPVVEQQVKWRRAREGFPQEMLDEE
MRKIAYSVRKLDDHLRDNEWLVPGMYTLADICNFAIANGMQHGYPELVNKEDTPGLLRWIEQINERPAAQKMFADVPREIRHEDAKK

pET28 LigF935 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGTTGACGCTGTACAGCTTTGGCCCTATG
GCAAATTCCTTGAAGCCTATGCTGACCCTGTTCGAAAAGTTTGAATTCGGTCGCGATTTTACCGTCCATCGTTTGGACCCGTCCAAANTTG
AGCATCATACGGATTGGTTCAAAGCAATTAACCCGCGTGGTCAAGTTCCGGCTCTGAAAGACGGCGATCGCATTGTGACCGAGAGCACC
GTCATTTGCGAATACTTGGAAGATGCTCACCCAACCGCCGTCAAGTTGCGTCCGGACGATCCGTACGATCGCGCGCAAATGCGTGTTTGG
ACCAAATGGGTGGATGAGTATTTCTGTTGGTGTGTGAGCACCATTGGTTGGCATCGTGGCGTTTCCCACATGGCCAAGGCGCTGAGCGA
CCAGGAATTCGAAGAACACTTGAAGAAGATTCCGATCCCGGAGCAACAGGTTAAGTGGCGTCGCGCGCGCGAAGGTTTCCCACAGGAT
TTGCTGGACGAGGAAATGCGTAAAATTGCGGTTAGCGTTCGCAAACTGGACGACCACCTGGCGGACAACGAGTGGCTGGCGGGTGGTA
TGTATTCCCTGGCGGACATCTGCAACTTCGCGATCGCGAATGGTATGCAAAATGGTTTCGCGGAGCTGGTTAACACCAGCGATACCCCGC
ACCTGGTGCGTTGGATCGAGCAGATCAACGCGCGTCCGAAAGTGCAACAGATGTTTGCGAGCGTGCCGCGTGAGCAACTGGGTCCGCC
GCGTTAAGCTT

MGSSHHHHHHSSGLVPRGSHMLTLYSFGPMANSLKPMLTLFEKFEFGRDFTVHRLDPSKXEHHTDWFKAINPRGQVPALKDGDRIVTESTVI
CEYLEDAHPTAVKLRPDDPYDRAQMRVWTKWVDEYFCWCVSTIGWHRGVSHMAKALSDQEFEEHLKKIPIPEQQVKWRRAREGFPQDLLD
EEMRKIAVSVRKLDDHLADNEWLAGGMYSLADICNFAIANGMQNGFAELVNTSDTPHLVRWIEQINARPKVQQMFASVPREQLGPPR

pET28 LigF965 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTGACCTTGTATTCCTTTGGTCCGGGT
GCGAATTCCCTGAAGCCACTGTTGACGCTGTACGAGAAGGGTCTGGAATTTAAAGGTGTTCGTTTGAATCCTGCCCAGTTTGAACACCAT
GAAGACTGGTTCAAAGCGATCAATCCTAATGGTCAGGTCCCGGCTCTGGTTGATAACGGTCGCGTGATTACGGAGTCCACCGTTATCTGC

GAGTACCTGGAGGATGAATATCCGACCGCAGTCAAATTGCGCCCGGCGGATAGCTACGGTCGCGCCCAGATGCGTGTGTGGACCAAGT
GGGTTGACGAATATTTCTGCTGGTGTGTGAGCACCATCGGCTGGAGCCGCATGATCAGCGGCATGGCTCGTGCTATTAGCGCAGAGGAA
TTTGAGGAGAAGATTAAACGCATTCCAATTCCGGAGCAGCAAGTGAAATGGCGTCGCGCACGCGACGGTTTTCCGCAAGACATGTTGGA
TGAGGAAATGCGTAAGATTGGCGTCAGCGTTCGTCGTCTGAACGATCATCTGGCCGACCACGAGTGGTTGGCGGGTGAAATGTACAGCT
TGGCGGATATCTGTAACTATGCGATTGCGGGTGGTATGCAATTCGGTTTCGCGGAGCTGGTTAATCAAGCGGATACCCCGCACCTGCTGC
GTTGGATCGAGATGATCGCGGCGCGTCCGGCGGCGAAGAAAATGGTTGCGGAAGTTCCGATGGAACGCCTGGTGGCGGACGCGTAAG
CTT

MGSSHHHHHHSSGLVPRGSHMLTLYSFGPGANSLKPLLTLYEKGLEFKGVRLNPAQFEHHEDWFKAINPNGQVPALVDNGRVITESTVICEYL
EDEYPTAVKLRPADSYGRAQMRVWTKWVDEYFCWCVSTIGWSRMISGMARAISAEEFEEKIKRIPIPEQQVKWRRARDGFPQDMLDEEMR
KIGVSVRRLNDHLADHEWLAGEMYSLADICNYAIAGGMQFGFAELVNQADTPHLLRWIEMIAARPAAKKMVAEVPMERLVADA

pET28 LigG NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCGGAGCCGCAAGAGCTGACGATCT
ACCACATCCCGGGTTGCCCGTTCAGCGAACGTGTGGAGATTATGCTGGAGCTGAAGGGTCTGCGTATGAAGGACGTTGAAATCGATATT
AGCAAGCCGCGTCCAGATTGGCTGCTGGCGAAAACGGGTGGTACGACCGCGCTGCCGCTGCTGGACGTTGAGAATGGTGAGAGCTTGA
AGGAAAGCATGGTGATTCTGCGTTACCTGGAGCAGCGTTACCCGGAGCCGGCGGTTGCACACCCGGATCCGTTCTGCCACGCGGTGGAA
GGTATGCTGGCGGAACTGGCGGGTCCATTCAGCGGTGCGGGTTACCGTATGATCCTGAACCGTGAGATTGGCAAACGTGAAGAAATGC
GTGCGGCGGTTGACGCGGAATTCGGTAAAGTGGATGCGTTTCTGAAACGTTACGCGACCGGTAGCGACTTCCTGTTCGATGATCGTTTC
GGTTGGGCGGAAGTTGCTTTCACCCCGATGTTCAAGCGTCTGTGGTTCCTGGATTACTACGAAGACTACGAGGTGCCGGCGAACTTCGA
CCGTGTTCTGCGTTGGCGTGCGGCATGTACGGCGCACCCAGCGGCGCAATACCGTAGCAAGGAAGAGCTGCTGAAACTGTACTACGACT
ACACCCAGGGTGGTGGCAACGGTCGTATCCCTGAGGGTCGTAGCATCAGCAGCTTCAGCCCGGACGTCGACTGGCGTACCCGTCCGATG
CCGCCGCGTGACAAGTGGGGTCATGCGGCGACCGACGCCGAACTGGGTCTGACCCGTTAAGCTT

MGSSHHHHHHSSGLVPRGSHMAEPQELTIYHIPGCPFSERVEIMLELKGLRMKDVEIDISKPRPDWLLAKTGGTTALPLLDVENGESLKESMVI
LRYLEQRYPEPAVAHPDPFCHAVEGMLAELAGPFSGAGYRMILNREIGKREEMRAAVDAEFGKVDAFLKRYATGSDFLFDDRFGWAEVAFTP
MFKRLWFLDYYEDYEVPANFDRVLRWRAACTAHPAAQYRSKEELLKLYYDYTQGGGNGRIPEGRSISSFSPDVDWRTRPMPPRDKWGHAAT
DAELGLTR

pET28 LigG-NS NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGGCGTGGCACGAGGAACCGCCGCCGG
GTGCGCTGCGTATGTACCACATTCCGGGCTGCCCGTTCAGCGAACGTGTGGAAATCCTGCTGGATCTGAAGGGTCTGAGCGGTATCATG
GACGATCACGAGGTGGACATCAGCAAGCCGCGTCCGGACTGGCTGCTGAGCAAGACCCGTGGCACCACCAGCCTGCCGGCGCTGGAAC
TGGAAAATGGTGAGACCCTGAAAGAGAGCATGGTTATTCTGCGTTACTTCGAAGAGCGTTTCCCGGAACGTCCGGTTGCGCAGCGTGAC
CCGTTCGAACACGCGGTTGAGGCGATGCTGTGCGCGACCGACGGTCAATTCACCGGTGCGGGTTACCGTATGATTCTGAATCGTGACCG
TGCGAAACGTGACGACTGCAAGGCGGAAGTTGACGCGCAATACGCGCGTCTGGACGACTTCCTGCGTCACTACAGCCCAGATGGTGACT
ACCTGTTCGCGAGCTTCGGTTGGGCGGAAGTGGTGTTTGCGCCGATGTTCAAACGTCTGTGGTTCCTGGAATACTACGAGGGTTATGAA
GTTCCACAGCACCTGGCGCGTGTGCTGCGTTGGCGTGAGGCGTGCACCGAGCACCCTGCGGTGGCGAACCGTCACAGCCACCGTGAGCT
GATGACCCTGTACTACGATTACAGCCAAGGTGGTGGTAACGGTCGTCTGCCGCAGGGTCGTCACATTAGCAGCTTCACCCTGGACCCGG
CGTGGGAAAAACGTCCGATGCCACCGCGTGACAAATGGGGCACCCCAGCAACCGATCGTGAACTGGGTCTGGTGACCGCGACCGTGGG
TGCGTAAGCTT

MGSSHHHHHHSSGLVPRGSHMAWHEEPPPGALRMYHIPGCPFSERVEILLDLKGLSGIMDDHEVDISKPRPDWLLSKTRGTTSLPALELENGE
TLKESMVILRYFEERFPERPVAQRDPFEHAVEAMLCATDGQFTGAGYRMILNRDRAKRDDCKAEVDAQYARLDDFLRHYSPDGDYLFASFGW
AEVVFAPMFKRLWFLEYYEGYEVPQHLARVLRWREACTEHPAVANRHSHRELMTLYYDYSQGGGNGRLPQGRHISSFTLDPAWEKRPMPPR
DKWGTPATDRELGLVTATVGA

pET28 LigG-TD NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGACCCGTCCGACCGTTTACCACATTCCG
GTCTGCCCGTTCTGCCAACGTGTTGAAATCCTGCTGAGCCTGAAAGGTCGTCGTGAGGATGTGGATTTCCGTATGATCGACATTACCGCG
CCGCGTCCGGACTGGCTGCTGCAAAAGACGCGTGGTACGACCGCACTGCCGGTGCTGGAGACCGCGGATGGTCGTGTTATTAAGGAAA
GCCTGGTGATCCTGCGTTACTTTGAAGATATCTACCGTGAGCCGCAAATTGCGCAAACCGATCCGTACCGTCGTGCGGTGGAGAATATGC
TGACCACCATGGACCGTGATTTCGTTGCGGCGGGTTACGGTTGGCTGATGAACCAAGATCCGAAACAACGTGATGCGCTGCGTGAAAAC
ATGCTGAAGCAGTACGCGCAGCTGGATGACTTCCTGCTGGAACACGGTGCGCCGGGTCCGTTCCTGTTCGAAACCTTCGGTTGGGCTGA
AACCGTTTTCACCCCGTTCTTCCAGCGTTTCTGGTTCCTGGAGTACTACGAGGATTTCGAACTGCCGACCGCGAGCCGTTACGCGCGTGTG
CGTGAATGGATCGATGCGTGCGTGGCGCACCCGGCCGCGCAACAGACCACCCGTGAAGAAGTGATTAAGCTGTACTATGATTACGCGTG

CGGTGCGGGTAATGGTGCGCTGCTGCCAGGTCGTAGCCGTAGCAGCTTCGCGCTGGAGCCGGATTGGCGTGCGCGTCCGTGGCCGCCG
CGTAGCAAGTACCGTCAACCGGCGGGTGACACCGAGCTGGGTCTGTAAGCTT

MGSSHHHHHHSSGLVPRGSHMTRPTVYHIPVCPFCQRVEILLSLKGRREDVDFRMIDITAPRPDWLLQKTRGTTALPVLETADGRVIKESLVILR
YFEDIYREPQIAQTDPYRRAVENMLTTMDRDFVAAGYGWLMNQDPKQRDALRENMLKQYAQLDDFLLEHGAPGPFLFETFGWAETVFTPFF
QRFWFLEYYEDFELPTASRYARVREWIDACVAHPAAQQTTREEVIKLYYDYACGAGNGALLPGRSRSSFALEPDWRARPWPPRSKYRQPAGD
TELGL

pET28 LigG817 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGCTGGAGCTGCTGACGAGCGAGACCCC
GAACGGCTGGAAGACCACGATTATGCTGGAGGAATTGGGTGTTGCCTACAAGGTGCGTCCTATTTCCTTGACGGATCGTGAGCAAAAGG
AAGAGTGGTACTTGGGCCTGAACCCGAATGGTCGTATCCCAACCTTGATCGACCACGACGTTGAGACCGAAGGCGGCTCCTTCGCAGTG
TTTGAAAGCGGTGCAATTTTGATTTATCTGGCCGAGAAGTTCGATCGTTTCCTGCCGAAATTCATCACCGGCCGTTCCGAGGCCATTCAGT
GGGTGATGTGGCAGATGAGCGGTTTGGGTCCGATGATGGGTCAGGCGACCGTCTTCAACCGTTACTTCGATGAGAAGCTGCCGAGCGT
TATTGACCGCTATACCCGCGAATCCCGTCGCCTGTTTGAAGTCATGGATCGCCGCTTGGCGGATCGCGAATTTCTGGCTGGCGACTACAG
CATTGCTGATATTGCCTGCTTCCCGTGGGTTCGTGGTCGTGACTGGGCCTGCATCGATATGGATGGCCTGTCCCATCTGCAACGTTGGTTC
GACACCGTGGCGGCGCGTCCTGCGGTGCAACGTGGTCTGCTGATCCCAGAGCCGCCGGCGGCGAGCGAAATGGCGGCGCAGACCACCC
GCCAGGGTAAGAATATCCTGGCCTAAGCTT

MGSSHHHHHHSSGLVPRGSHMLELLTSETPNGWKTTIMLEELGVAYKVRPISLTDREQKEEWYLGLNPNGRIPTLIDHDVETEGGSFAVFESG
AILIYLAEKFDRFLPKFITGRSEAIQWVMWQMSGLGPMMGQATVFNRYFDEKLPSVIDRYTRESRRLFEVMDRRLADREFLAGDYSIADIACFP
WVRGRDWACIDMDGLSHLQRWFDTVAARPAVQRGLLIPEPPAASEMAAQTTRQGKNILA

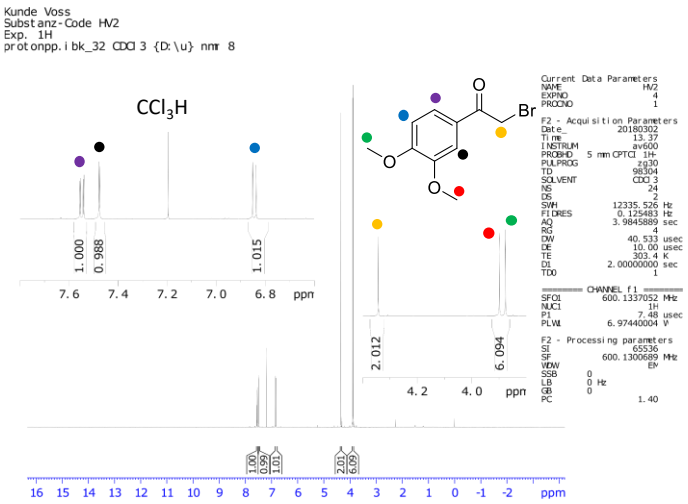
pET28 GST3 NdeI HindIII

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATGTTGGAGCTGTGGACGAGCGAGACCC
CGAACGGTTGGAAGACCACCATTATGTTGGAGGAATTGGATGCGAATTACACCCTGCGCCCGATTAGCCTGACCAATCGCGAACAGAAG
GAAGACTGGTATCTGGCGCGTAACCCGAACGGTCGCATCCCGACCCTGATCGACCACGAAGTGGATGCGGGTAATGGTGGCTTTGCGGT
GTTCGAAAGCGGCGCGATCTTGATTTACTTGGCGGAAAAGTTCGGTCGTTTTCTGCCGGCCGATACGATGGGCCGCAGCCGCGCGATTC
AGTGGGTTATGTGGCAGATGTCCGGTTTGGGTCCGATGATGGGTCAAGCGACCGTCTTCAACCGTTACTTTGAGCCGCGTTTGCCAGAG
GTGATCGACCGTTACACGCGCGAATCCCGCCGTCTGTTCGAGGTTATGGATACCCATCTGGCAGACAACGAATTTCTGGCGGGTGACTAC
AGCATCGCGGATATCGCGTGTTTCCCGTGGGTTCGTGGTCATGACTGGGCATGCATCGATATGGAAGGTCTGCCTCACCTGCAACGCTG
GTTTGAGACGATTGGTGAACGCCCGGCTGTTCAACGTGGTTTGCTGCTGCCTGAGCCGCCGAAAGCGGATGAGATGGCGGAGAAGACC
ACGCGTCAAGGTAAGAACATTCTGGCGTAAGCTT

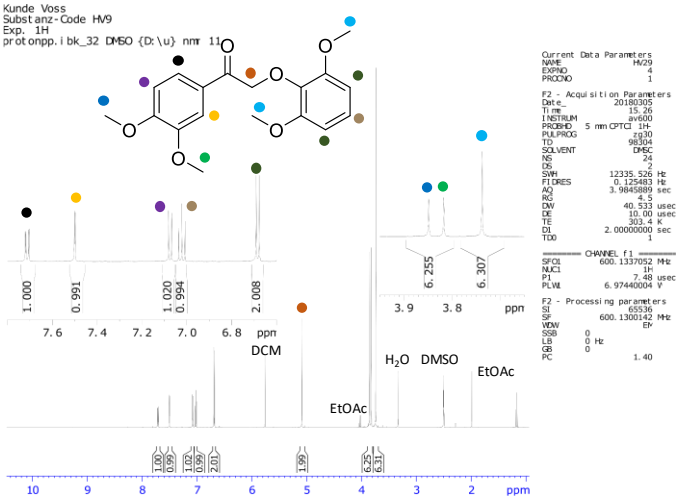
MGSSHHHHHHSSGLVPRGSHMLELWTSETPNGWKTTIMLEELDANYTLRPISLTNREQKEDWYLARNPNGRIPTLIDHEVDAGNGGFAVFES
GAILIYLAEKFGRFLPADTMGRSRAIQWVMWQMSGLGPMMGQATVFNRYFEPRLPEVIDRYTRESRRLFEVMDTHLADNEFLAGDYSIADIA
CFPWVRGHDWACIDMEGLPHLQRWFETIGERPAVQRGLLLPEPPKADEMAEKTTRQGKNILA

6.2 NMR Spectra

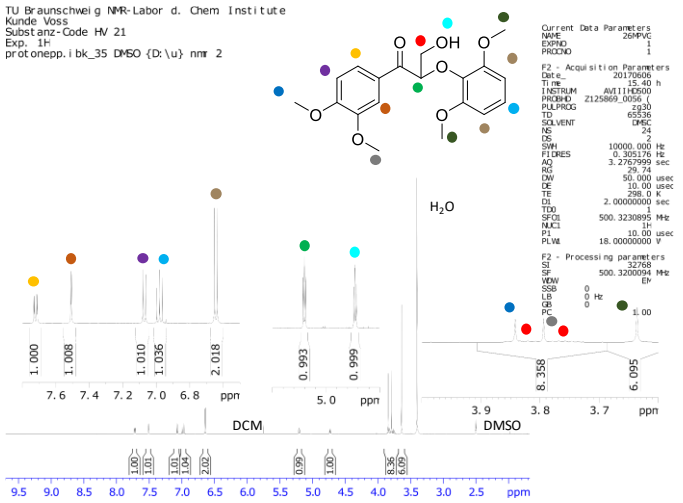
Compound **1**



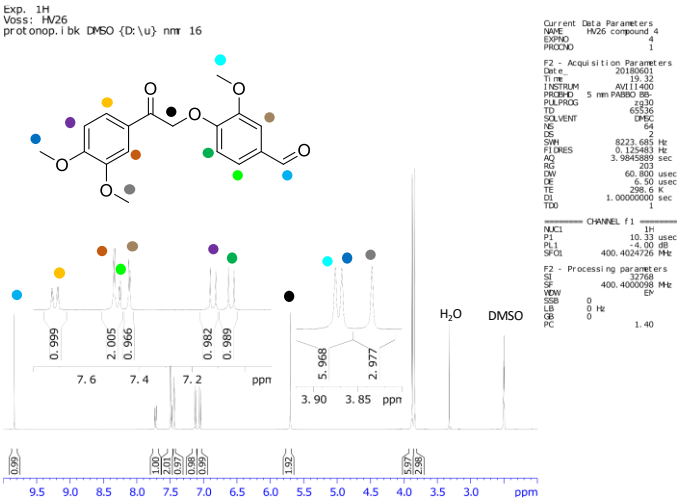
Compound **2**



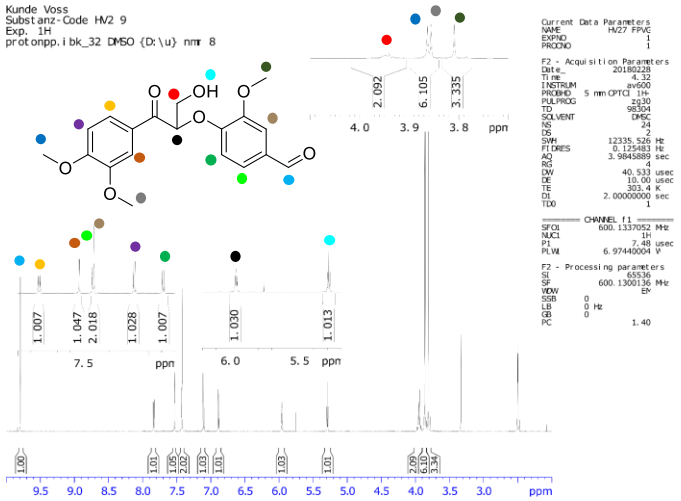
Compound **3**



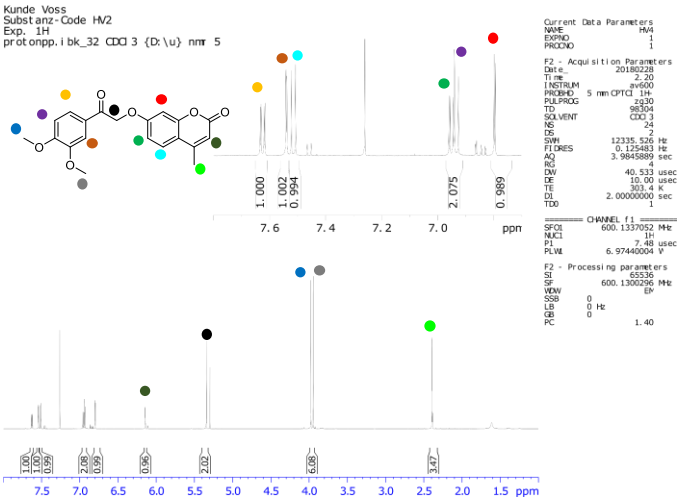
Compound **4**



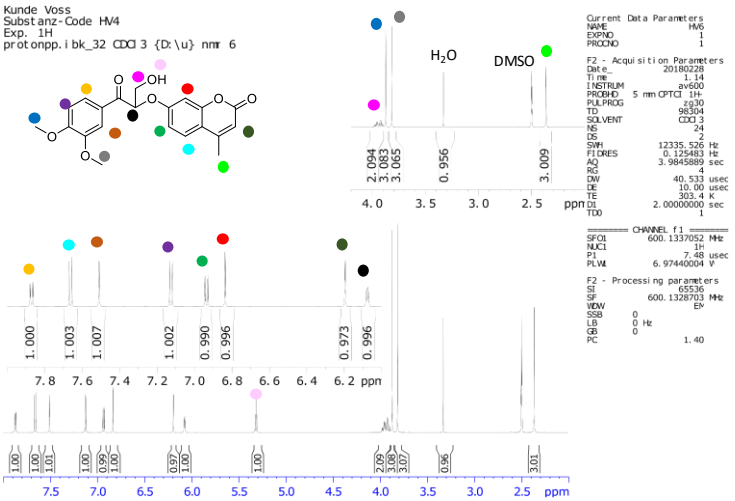
Compound **5**



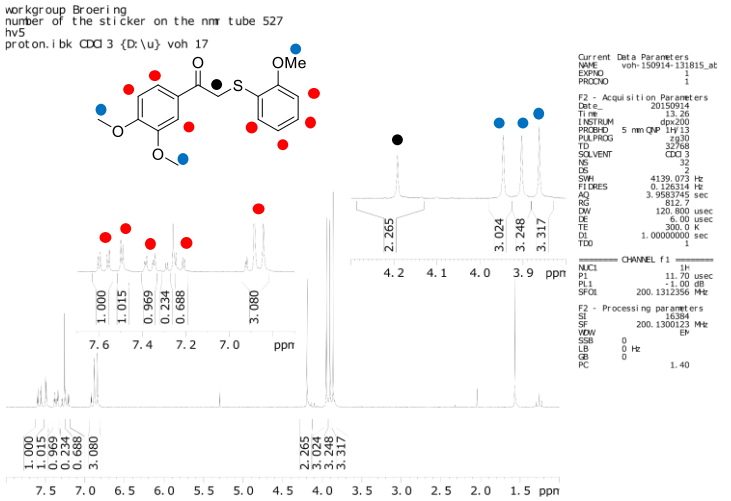
Compound **6**



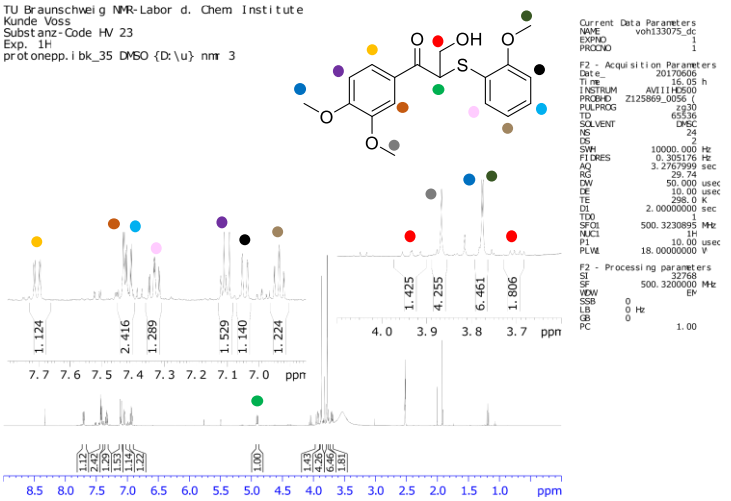
Compound **7**



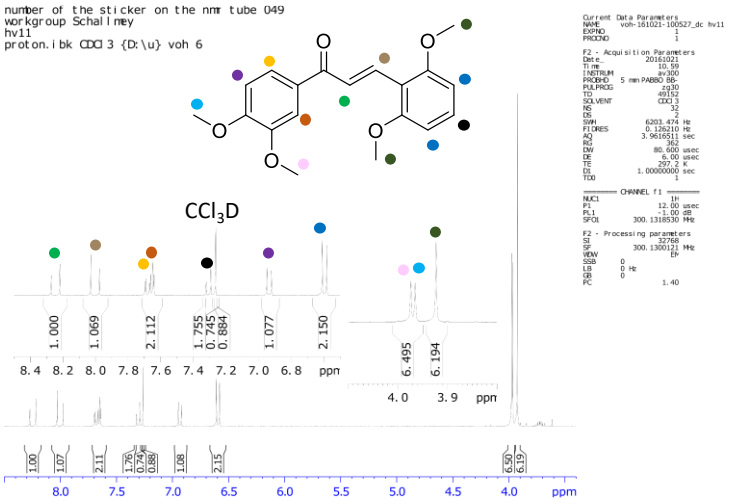
Compound **8**



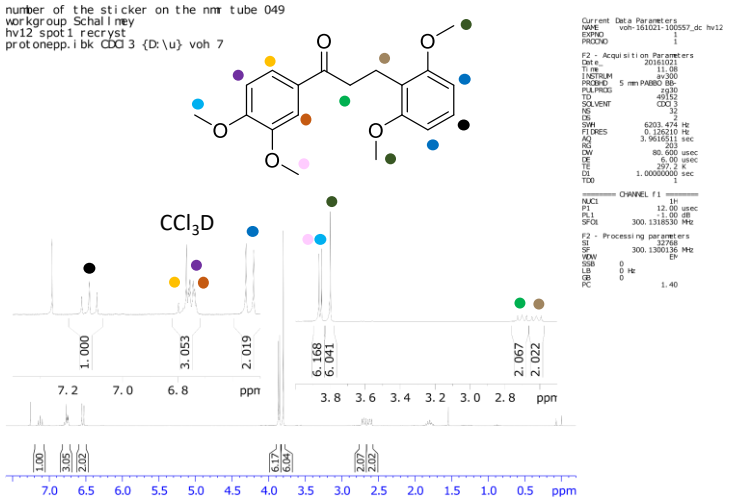
Compound **9**



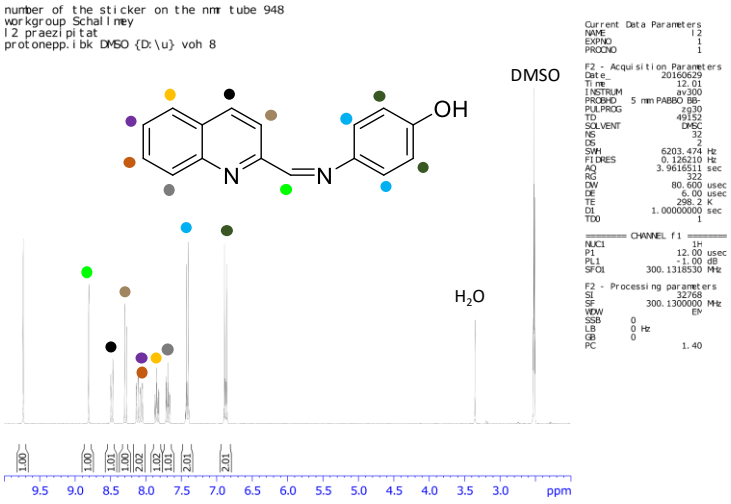
Compound **10**



Compound **11**



Compound **12**



6.3 HPLC-chromatograms

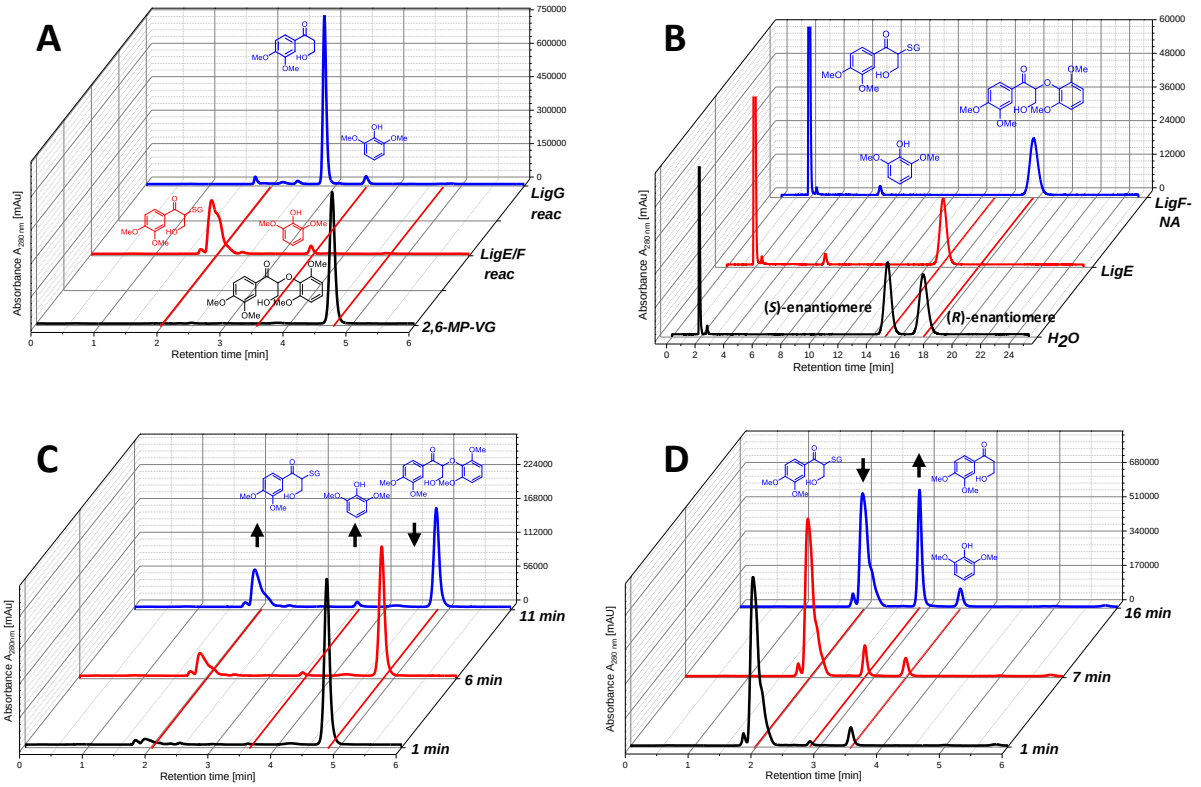


Figure 6.1 HPLC chromatograms of β-etherase and glutathione reactions. **A** typical HPLC chromatograms of before and after β-etherase and glutathione reaction. **B** typical chiral HPLC chromatograms before and after β-etherase reaction. **C** and **D** typical reaction progress of β-etherase and glutathione reaction analyzed with HPLC.

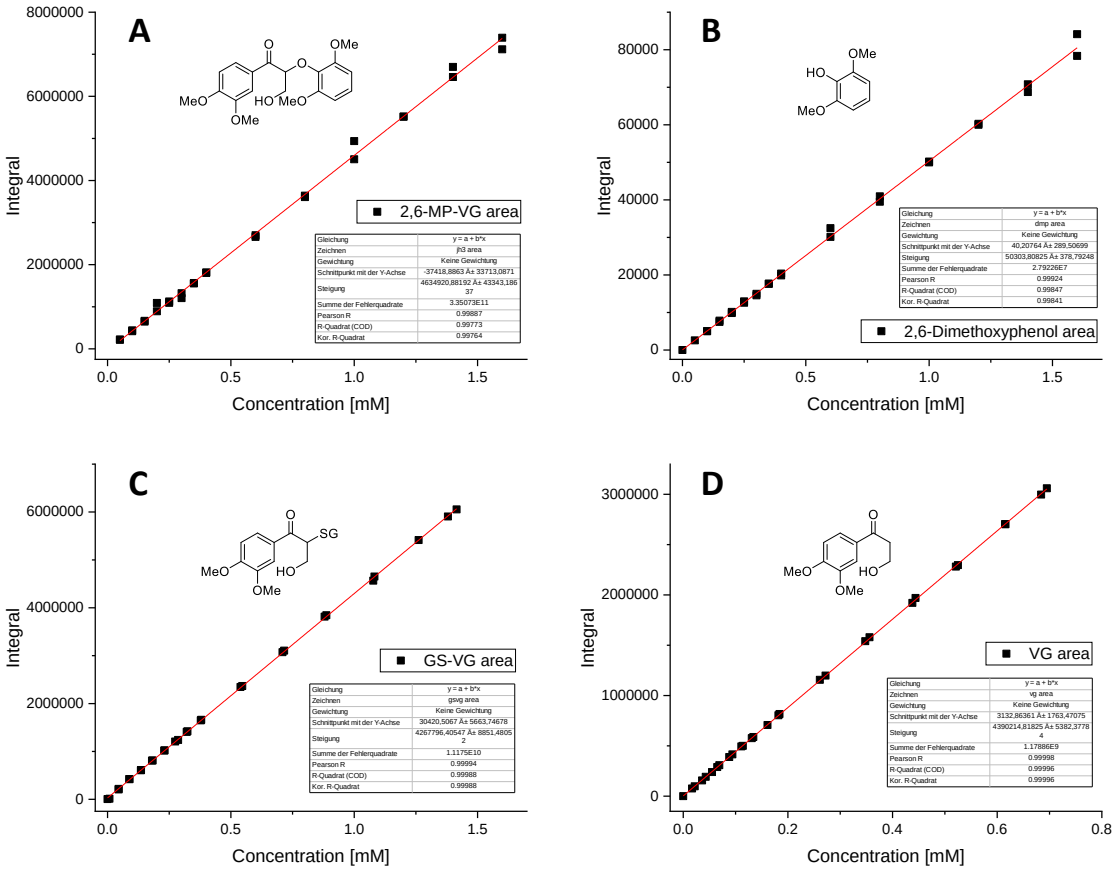


Figure 6.2 HPLC calibration curves. **A** 2,6-MP-VG, **B** 2,6-dimethoxypenol, **C** GS-VG, **D** VG

6.4 Sequences grouped by the PPR algorithm and used for database mining

GenBank accession numbers of enzymes used for database mining by PPR or phylogenetic analysis.

β-etherases			
Gene	GenBank accession number	Gene	GenBank accession number
LigE	WP_014075192	LigF	WP_014075191
LigE-NA	WP_011446047	LigF-NA	WP_041551020
LigE-NS	WP_013832481	LigF-NS	WP_013832480
GST5	WP_039391125	NaLigF1	ABD26530
LigP	WP_014077574	NaLigF2	ABD27301
LigE179	WP_046903179	GST4	WP_039391123
LigE283	OJU60283	LigF008	WP_055919008
LigE491	WP_044331491	LigF215	OGT78215
LigE760	ODU84760	LigF729	ODU83729
LigE889	WP_055920889	LigF755	WP_066854755
LigE915	WP_062781915	LigF921	WP_054529921
		LigF935	OJU59935
		LigF965	WP_068075965

Glutathione lyasen			
Gene	GenBank accession number	Gene	GenBank accession number
LigG	BAA77216	LigG817	WP_041558817
LigG-NS	CCA92089	GST3	WP_039391121
LigG-TD	WP_011311562	$NaGST_{Nu}$	WP_011446237
GST6	GAM05532	$SYKGST_{Nu}$	BAK65087

GenBank accession numbers of protein sequences grouped together with LigE, LigF, NaLigF2 and LigG in the PPR run. Some proteins are sharing the exact same protein sequence, in this case only one protein with this sequence was used for the phylogenetic tree (Figure 3.4). The GenBank of the excluded genes are marked in brackets.

LigE homologs (BLASTP queries: LigE, LigE-NA, LigE-NS, LigP, GST5)			
OYW49013	OYX94051	OYX66856	WP_014075192
WP_062781915	WP_054530882	OJW69813	WP_109797409
WP_107715663	WP_046903062	OJY67521	(ALL97524)
WP_054131978	WP_054106368	WP_054530343	(pdb|4YAM|A)
WP_028657990	BBC73886	WP_055921561	WP_039391125
WP_011446047	OJW72302	ODT90752	WP_044331491
WP_067913477	OJU60283	WP_014077574	WP_054436034
OYU35660	ODU67585	WP_088311142	WP_110873471
WP_062343074	GAO55515	WP_067743869	CDO38110
WP_039335072	WP_055920889	ODU84760	WP_054947729
WP_054121822	WP_104830666	WP_028641482	WP_079730390
WP_072379148	OYX64790	WP_046903179	WP_013832481
WP_084280065	OYW45403	WP_067615430	OJY68654
Outgroup: WP_011311562			
LigF homologs (BLASTP queries: LigF, LigF-NA, LigF-NS, NaLigF-1, GST4, NaLigF-2)			
WP_067910024	(KHS49048)	WP_055919008	BBC73091
ODU70795	WP_082356151	WP_107124580	WP_039391123
OYU33368	(SMC30538)	WP_088182463	OJU17901
WP_022675760	WP_082734789	WP_017182419	PVX29992
WP_072381855	OYW45183	WP_019052363	WP_088309375
WP_062782323	ODU83729	WP_046904735	(pdb|4XT0|A)
WP_054121580	OYX94961	WP_056685155	WP_014075191
WP_107715567	OYW49868	WP_066818898	(ALL97521)
GAO55900	OJW61742	WP_066804606	WP_068075965
OYX60692	OJU59935	BBD02394	WP_013832480
KPF91204	WP_054529921	WP_066609183	WP_054436033
WP_084276739	OZA57515	WP_100283311	OJY68655
WP_041551020	WP_066854755	WP_104831261	WP_054947728
ABD26530	WP_093081587	WP_044331490	WP_110873409
WP_082013229	WP_093018627	(OYX66589)	OGT78215
Outgroup: WP_011311562			
NaLigF-2 homologs (BLASTP queries: LigF, LigF-NA, LigF-NS, NaLigF-1, GST4, NaLigF-2)			
WP_052241956	WP_107716966	PIX64606	OZA20480

ABD27301	WP_082697614	WP_082347938	OLB14823
GAO54720	KUR73010	ODT86087	OYZ46457
WP_083760905	WP_054106614	OJY66881	WP_088182917
WP_084279863	OJU60146	WP_067198479	WP_066818877
SMC83186	WP_068074729	WP_082703412	WP_051280663
WP_055925679	WP_072381484	KUR79160	OZA31394
ODU69502	ODU70978	WP_081912041	OGT83286
OYX94580	WP_088308200	WP_100283320	WP_054106711
OYX62548	OJW70795	WP_067615204	OYZ28452
OYU33440	WP_054531108	WP_091149001	PZN34063
WP_054132647	WP_022677625	SFG44676	BBC71095
WP_062787547	OYW44946	WP_084238447	WP_110872575
OYW48665			
LigG homologs (BLASTP queries: LigG, LigG-NS, LigG-TD, GST6)			
WP_119744654	SER63208	WP_130754438	ODU85154
WP_054436063	WP_054947751	WP_120253484	WP_116470882
TCM36116	WP_093018625	WP_100284490	OJY68724
AJP74359	OJU18076	GAM05532	WP_041392591
SEI61191	WP_066854757	WP_052322378	ALL97522
WP_093081585	WP_041558818	WP_110873470	pdb|4YAP|A
WP_107124579	WP_017182418	WP_066609229	WP_114952997
WP_019052364	CCA92089	KMS52183	TAJ96034
CDO38111	WP_066819246	WP_066804683	pdb|4G10|A
SLJ97389	WP_056685274	WP_088182513	pdb|4YAV|A
WP_079730391	GBH32489		

6.5 Glutathione reductase screening libaries

Results of the relative activity screening of the LigG-TD libraries using the glutathione reductase assay. The mutants were expressed in LB media in microtiter plates. The cell pellets were lysed with B-PER and the assay was performed with GS-VG as substrate and NADPH and glutathione reductase for detection of the glutathione lyase reaction. The NADPH decrease was monitored at 360_{nm}. The slope was determined, normalized at OD_{600} and the difference to the mean value of the wild-type controls in percent calculated (section 2.3.7.4). Wells with without cell growth are marked with NG. Wildtype controls are highlighted in red, empty vector controls in green and mutants chosen for sequencing in bold.

V11X (NDT+VMA+ATG+TGG)

	1	2	3	4	5	6	7	8	9	10	11	12
A	-29	-66	-39	-40	-20	-27	-37	-18	-56	62	42	49
B	-65	-22	-85	-47	61	-37	48	105	-7	8	-37	109
C	-51	-55	-41	-15	-35	42	75	-13	129	105	85	9
D	-52	-66	-76	-64	-66	-48	-55	-49	-47	-45	13	52
E	-56	-51	-52	-24	1	-28	-19	4	-16	-6	-18	53
F	-54	-64	-57	-5	-53	-60	-24	7	69	-10	-42	35
G	-30	-44	-81	-45	-53	-46	-31	-52	-19	-36	-11	-3
H	-70	-73	-36	-24	-51	-14	11	-69	-50	53	29	-12

A55X (NDT+VMA+ATG+TGG)

	1	2	3	4	5	6	7	8	9	10	11	12
A	-41	32	-49	-58	-5	-44	-51	-57	-58	-50	-51	-8
B	-65	12	8	-38	-32	-28	-47	-18	16	-47	-52	NG
C	-48	47	-7	16	6	11	-20	-18	-21	-39	-41	-40
D	-53	-51	46	-40	-46	35	-2	-40	7	-41	-20	-53
E	-64	-40	-51	-54	114	-74	-30	-56	62	11	-51	-50
F	-67	-52	-37	26	-10	12	-18	28	-52	-39	-67	-10
G	-12	-26	-44	-21	-52	-33	-22	-48	32	-29	-37	-8
H	-26	-47	-43	-11	-38	-65	39	-26	34	-45	-61	-23

F165X (NDT+VMA+ATG+TGG) No. 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	23	-26	23	-34	46	-14	48	-28	-54	28	-18	14
B	-17	-11	65	-65	-64	-69	70	53	30	-17	-71	62
C	22	-27	25	-59	-100	-100	0	29	-61	-70	-100	-97
D	-2	-72	-100	-70	-68	-74	-63	-72	-57	-59	-85	43
E	-65	-100	-60	-32	111	-38	-51	-76	-23	-35	-70	-76
F	-55	-100	-100	-100	-6	-85	-85	-68	9	-94	-100	-87
G	-66	-12	10	-63	-19	-74	-9	-34	-81	-67	-7	-71
H	-71	-73	-69	-47	-71	-69	-64	-69	-55	-49	-19	-68

F165X (NDT+VMA+ATG+TGG) No. 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	15	-100	-9	12	8	-8	-11	12	5	3	-2	-24
B	8	34	49	22	21	-1	-3	3	1	6	30	29
C	-20	-15	16	-100	2	37	1	-100	25	-4	6	7
D	-3	-16	-38	-19	-8	3	-17	-7	-3	-25	-100	13
E	3	6	-29	-6	-36	-35	39	-19	-14	-52	-30	-10
F	-13	-26	5	-10	-14	-25	-28	-35	-31	3	-33	12
G	-11	-17	-41	-14	-36	-15	-32	-43	-22	-30	-23	-11
H	7	-13	-16	-8	15	0	3	-8	6	-5	35	-100

V108X (NNS) No. 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	-4	-5	-31	-27	-30	-31	2	-15	-5	-10	-37	0
B	-24	-1	17	3	14	-38	21	18	-20	-46	-40	4
C	-29	13	-40	-10	-28	-25	-3	-17	15	-47	-32	-2
D	22	24	2	12	10	-15	17	3	9	-46	37	-34
E	16	13	8	8	8	17	5	-8	4	-1	-32	6
F	17	-25	-23	10	-4	1	-27	-13	9	11	-11	13
G	38	-43	36	11	-8	-4	27	7	-1	-2	-15	-15
H	-3	-25	-8	-33	-5	9	-25	-34	-42	-5	-4	2

V108X (NNS) No. 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	-5	11	-12	15	-3	-23	-29	-11	-23	-10	8	14
B	-14	-3	-18	-3	-10	-6	-41	-34	6	-22	-31	-26
C	8	4	-44	19	4	10	-53	-7	15	23	-37	25
D	8	6	-43	-6	7	-6	3	12	-25	11	7	24
E	-40	11	13	7	22	28	1	15	22	25	12	16
F	-20	2	-31	13	15	-45	-16	-29	16	-5	21	33
G	0	-1	-1	-2	11	-7	0	-37	-30	-4	3	18
H	3	1	0	-16	16	12	-17	-5	-23	-22	3	-4

M116 (NNS) No. 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	-5	-1	0	-34	3	6	-35	-10	18	-15	1	5
B	-2	1	12	1	-10	0	-25	-32	16	9	-13	-26
C	-35	-5	-2	3	-35	-11	-42	-15	-7	12	6	11
D	-1	5	-20	11	-5	-12	-32	-18	22	15	22	9
E	7	21	12	14	-12	9	1	5	23	27	14	13
F	-20	10	0	1	3	-2	-12	-20	-25	22	2	19
G	9	16	-2	8	-8	-13	-1	-22	1	-13	2	-7
H	1	-21	6	-1	-38	-26	-6	-1	-4	-5	1	2

M116 (NNS) No. 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	-6	-10	14	8	-21	4	9	5	19	-7	18	-1
B	-36	-31	-14	4	-19	-29	3	-33	15	12	-4	-15
C	-18	-5	20	5	-5	-1	-20	11	22	-1	10	-3
D	-15	-10	-7	6	1	18	8	-32	-2	21	3	-37
E	3	-4	-10	-6	-15	21	-5	0	-27	27	-3	13
F	10	-18	30	-29	-31	3	-8	10	11	25	0	25
G	-27	-4	-5	-9	-20	-7	-18	5	-7	6	-5	24
H	3	-18	-2	-9	-9	24	-4	-7	12	8	4	8

N223X (NNS) No. 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	7	-12	-24	4	-31	7	-25	-26	-33	-17	-27	-36
B	-27	-30	-5	-2	4	-20	-11	0	-36	-7	-34	-2
C	-15	-35	-29	-25	-31	-24	-5	-42	-3	-6	0	2
D	28	0	-37	-26	3	6	-34	12	-27	-26	-33	27
E	-25	-2	-27	12	13	11	-3	11	-29	-35	-7	-37
F	-20	-27	-23	-23	-27	1	-2	12	4	3	-39	5
G	-3	-34	-4	-38	-38	-10	-26	71	-16	-4	-13	-41
H	-20	1	-6	-3	5	5	-25	-29	-19	-26	11	-16

N223X (NNS) No. 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	**3**	**12**	-30	-8	-31	-30	-1	0	-15	-6	-31	-1
B	-8	-2	**-12**	**-23**	-17	-6	-4	-3	-33	-15	-3	**28**
C	-43	-50	-47	-33	-50	14	4	-47	-34	-31	-35	-31
D	6	-21	-2	7	**4**	**-9**	-24	-26	-23	-29	-35	6
E	-39	18	-39	-12	-22	2	**5**	**-3**	7	-3	-28	**34**
F	-24	-24	-21	-32	-18	11	-30	-29	9	13	**30**	20
G	6	10	-2	-31	14	9	11	-27	-19	-20	-13	-6
H	-31	**30**	-6	-18	-4	-10	19	-6	12	12	**1**	**22**

References

[1] Faunhofer ISE, “Stromerzeugung | Energy Charts,” can be found under https://www.energy-charts.de/energy_pie_de.htm, **2018**.

[2] Bundesverband der deutschen Bioethanolwirtschaft e.V. BDB, “Bioethanol Marktdaten: Bioethanolverbrauch, Bioethanolverwendung, Bioethanol-Rohstoffeinsatz, Bioethanolproduktion, Biosprit E10,” can be found under https://www.bdbe.de/daten/marktdaten-deutschland, **2018**.

[3] Bundesministerium für Ernährung und Landwirtschaft BMEL, “BMEL - Fragen und Antworten zu E10,” can be found under https://www.bmel.de/DE/Landwirtschaft/Nachwachsende-Rohstoffe/Bioenergie/e10/FAQ_node.html, **2018**.

[4] P. N. R. Vennestrøm, C. M. Osmundsen, C. H. Christensen, E. Taarning, *Angew. Chemie Int. Ed.* **2011**, *50*, 10502–10509.

[5] F. Souto, V. Calado, N. Pereira, *Mater. Res. Express* **2018**, *5*, 072001.

[6] A. T. W. M. Hendriks, G. Zeeman, *Bioresour. Technol.* **2009**, *100*, 10–18.

[7] K. Iiyama, T. Lam, B. A. Stone, *Plant Physiol.* **1994**, *104*, 315–320.

[8] J. Zakzeski, P. C. A. Bruijnincx, A. L. Jongerius, B. M. Weckhuysen, *Chem. Rev.* **2010**, *110*, 3552–3599.

[9] E. Sjostrom, *Wood Chemistry: Fundamentals and Applications*, **1993**.

[10] T. D. Bugg, M. Ahmad, E. M. Hardiman, R. Singh, *Curr. Opin. Biotechnol.* **2011**, *22*, 394–400.

[11] J. Ralph, K. Lundquist, G. Brunow, F. Lu, H. Kim, P. F. Schatz, J. M. Marita, R. D. Hatfield, S. A. Ralph, J. H. Christensen, et al., *Phytochem. Rev.* **2004**, *3*, 29–60.

[12] R. Rinaldi, R. Jastrzebski, M. T. Clough, J. Ralph, M. Kennema, P. C. A. Bruijnincx, B. M. Weckhuysen, *Angew. Chemie Int. Ed.* **2016**, *55*, 8164–8215.

[13] F. H. Isikgor, C. R. Becer, *Polym. Chem.* **2015**, *6*, 4497–4559.

[14] C. European comission, “LIBERATE,” can be found under https://projects.leitat.org/liberate/, **2019**.

[15] C. European comission, “Bacterial Enzymes and Bioprocesses for Lignin Valorisation | Projects | H2020 | CORDIS | European Commission,” **2019**.

[16] E. Adler, *Wood Sci. Technol.* **1977**, *11*, 169–218.

[17] O. Faix, *Holzforschung* **1991**, *45*, 21–28.

[18] X. Tian, Z. Fang, R. L. Smith, Z. Wu, M. Liu, Springer, Singapore, **2016**, pp. 3–33.

[19] P. Picart, P. D. De María, A. Schallmey, *Front. Microbiol.* **2015**, *6*, 1–8.

[20] P. Gallezot, *Catal. Today* **2007**, *121*, 76–91.

[21] J. E. Holladay, J. F. White, J. J. Bozell, D. Johnson, *PNNL* **2007**, *II*.

[22] G. A. Smook, *J. Chem. Technol. Biotechnol.* **2007**, *45*, 15–27.

[23] F. S. Chakar, A. J. Ragauskas, *Ind. Crops Prod.* **2004**, *20*, 131–141.

[24] Z. Fang, R. L. Smith, *Production of Platform Chemicals from Sustainable Resources*, Springer Singapore, Singapore, **2017**.

[25] D. Weidener, H. Klose, W. Leitner, U. Schurr, B. Usadel, P. Domínguez de María, P. M. Grande, *ChemSusChem* **2018**, *11*, 2051–2056.

[26] Tsutomu Ikeda, Kevin Holtman, John F. Kadla, A. Hou-min Chang, H. Jameel, *J. Agric. Food Chem.* **2002**,

50, 129–135.

[27] D. Wu, S.; Argyropoulos, *J. Pulp Pap. Sci* **2003**, *29*, 235–240.

[28] Z. Sun, B. Fridrich, A. de Santi, S. Elangovan, K. Barta, *Chem. Rev.* **2018**, *118*, 614–678.

[29] Anderson Guerra, Ilari Filpponen, A. Lucian A. Lucia, D. S. Argyropoulos, *J. Agric. Food Chem.* **2006**, *54*, 9696–9705.

[30] P. Picart, C. Müller, J. Mottweiler, L. Wiermans, C. Bolm, P. D. De Marfa, A. Schallmey, *ChemSusChem* **2014**, *7*, 3164–3171.

[31] D. L. Gall, J. Ralph, T. J. Donohue, D. R. Noguera, *Environ. Sci. Technol.* **2014**, *48*, 12454–12463.

[32] D. L. Gall, J. Ralph, T. J. Donohue, D. R. Noguera, *Curr. Opin. Biotechnol.* **2017**, *45*, 120–126.

[33] C. S. Lancefield, O. S. Ojo, F. Tran, N. J. Westwood, *Angew. Chemie - Int. Ed.* **2015**, *54*, 258–262.

[34] P. Picart, H. Liu, P. M. Grande, N. Anders, L. Zhu, J. Klankermayer, W. Leitner, P. Domínguez de María, U. Schwaneberg, A. Schallmey, *Appl. Microbiol. Biotechnol.* **2017**, *101*, 6277–6287.

[35] A. P. M. Tavares, B. Pinho, O. Rodriguez, E. A. Macedo, *Procedia Eng.* **2012**, *42*, 226–230.

[36] H. Liu, L. Zhu, M. Bocola, N. Chen, A. C. Spiess, U. Schwaneberg, *Green Chem.* **2013**, *15*, 1348.

[37] L. Devi, K. J. Ptasinski, F. J. J. . Janssen, *Biomass and Bioenergy* **2003**, *24*, 125–140.

[38] L. Fan, Y. Zhang, S. Liu, N. Zhou, P. Chen, Y. Cheng, M. Addy, Q. Lu, M. M. Omar, Y. Liu, et al., *Bioresour. Technol.* **2017**, *241*, 1118–1126.

[39] P. J. de Wild, W. J. J. Huijgen, H. J. Heeres, *J. Anal. Appl. Pyrolysis* **2012**, *93*, 95–103.

[40] K. Barta, G. R. Warner, E. S. Beach, P. T. Anastas, *Green Chem.* **2014**, *16*, 191–196.

[41] L. Pollegioni, F. Tonin, E. Rosini, *FEBS J.* **2015**, *282*, 1190–1213.

[42] M. F. Hullo, I. Moszer, A. Danchin, I. Martin-Verstraete, *J. Bacteriol.* **2001**, *183*, 5426–30.

[43] G. Janusz, A. Pawlik, J. Sulej, U. Swiderska-Burek, A. Jarosz-Wilkolazka, A. Paszczynski, *FEMS Microbiol. Rev.* **2017**, *41*, 941–962.

[44] P. J. Teunissen, J. A. Field, *FEBS Lett.* **1998**, *439*, 219–23.

[45] V. Christian, R. Shrivastava, D. Shukla, H. Modi, B. R. M. Vyas, *Enzyme Microb. Technol.* **2005**, *36*, 327–332.

[46] M. Fabbrini, C. Galli, P. Gentili, *J. Mol. Catal. B Enzym.* **2002**, *16*, 231–240.

[47] P. K. Chaurasia, S. Yadava, S. L. Bharati, S. K. Singh, *Green Chem. Lett. Rev.* **2014**, *7*, 100–104.

[48] A. E. Lohning, A. E. Salinas, *Curr. Med. Chem.* **1999**, *6*, 279–309.

[49] C. Ioannides, *Enzym. Syst. that Metab. Drugs Other Xenobiotics* **2002**, *4*, 1–32.

[50] E. Masai, S. Nishikawa, N. Morohoshi, T. Haraguchi, *FEBS Lett.* **1989**, *249*, 348–352.

[51] E. Masai, Y. Katayama, S. Kawai, S. Nishikawa, M. Yamasaki, N. Morohoshi, *J. Bacteriol.* **1991**, *173*, 7950–7955.

[52] E. Masai, Y. Katayama, S. Kubota, S. Kawai, M. Yamasaki, N. Morohoshi, *FEBS Lett.* **1993**, *323*, 135–140.

[53] E. Masai, S. Kubota, Y. Katayama, S. Kawai, M. Yamasaki, N. Morohoshi, *Biosci. Biotechnol. Biochem.* **1993**, *57*, 1655–1659.

[54] E. Masai, Y. Katayama, S. Nishikawa, M. Fukuda, *J. Ind. Microbiol. Biotechnol.* **1999**, *23*, 364–373.

[55] E. Masai, A. Ichimura, Y. Sato, K. Miyauchi, Y. Katayama, M. Fukuda, *J. Bacteriol.* **2003**, *185*, 1768–1775.

[56] D. Kasai, E. Masai, Y. Katayama, M. Fukuda, *FEMS Microbiol. Lett.* **2007**, *274*, 323–328.

[57] Y. Sato, H. Moriuchi, S. Hishiyama, Y. Otsuka, K. Oshima, D. Kasai, M. Nakamura, S. Ohara, Y. Katayama, M. Fukuda, et al., *Appl. Environ. Microbiol.* **2009**, *75*, 5195–5201.

[58] D. L. Gall, H. Kim, F. Lu, T. J. Donohue, D. R. Noguera, J. Ralph, *J. Biol. Chem.* **2014**, *289*, 8656–8667.

[59] P. Picart, M. Sevenich, P. Domínguez de María, A. Schallmey, *Green Chem.* **2015**, *17*, 4931–4940.

[60] W. S. Kontur, C. A. Bingman, C. N. Olmsted, D. R. Wassarman, A. Ulbrich, D. L. Gall, R. W. Smith, L. M. Yusko, B. G. Fox, D. R. Noguera, et al., *J. Biol. Chem.* **2018**, *293*, 4955–4968.

[61] W. S. Kontur, C. N. Olmsted, L. M. Yusko, A. V Niles, K. A. Walters, E. T. Beebe, K. A. Vander Meulen, S. D. Karlen, D. L. Gall, D. R. Noguera, et al., *J. Biol. Chem.* **2019**, *294*, 1877–1890.

[62] J. Husarcíková, H. Voß, P. Domínguez de María, A. Schallmey, *Appl. Microbiol. Biotechnol.* **2018**, *102*, 5391–5401.

[63] Y. Ohta, R. Hasegawa, K. Kurosawa, A. H. Maeda, T. Koizumi, H. Nishimura, H. Okada, C. Qu, K. Saito, T. Watanabe, et al., *ChemSusChem* **2017**, *10*, 425–433.

[64] D. L. Gall, W. S. Kontur, W. Lan, H. Kim, Y. Li, J. Ralph, T. J. Donohue, D. R. Noguera, *Appl. Environ. Microbiol.* **2018**, *84*, e02076-17.

[65] J. Reiter, H. Strittmatter, L. O. Wiemann, D. Schieder, V. Sieber, *Green Chem* **2013**, *15*, 1373–1381.

[66] M. Marinović, P. Nousiainen, A. Dilokpimol, J. Kontro, R. Moore, J. Sipilä, R. P. de Vries, M. R. Mäkelä, K. Hildén, *ACS Sustain. Chem. Eng.* **2018**, *6*, 2878–2882.

[67] Y. Otsuka, T. Sonoki, S. Ikeda, S. Kajita, M. Nakamura, Y. Katayama, *Eur. J. Biochem.* **2003**, *270*, 2353–2362.

[68] Y. Ohta, S. Nishi, R. Hasegawa, Y. Hatada, *Sci. Rep.* **2015**, *5*, 15105.

[69] K. E. Helmich, J. H. Pereira, D. L. Gall, R. A. Heins, R. P. McAndrew, C. Bingman, K. Deng, K. C. Holland, D. R. Noguera, B. A. Simmons, et al., *J. Biol. Chem.* **2016**, *291*, 5234–5246.

[70] E. Meux, P. Prosper, E. Masai, G. Mulliert, S. Dumarçay, M. Morel, C. Didierjean, E. Gelhaye, F. Favier, *FEBS Lett.* **2012**, *586*, 3944–3950.

[71] J. H. Pereira, R. A. Heins, D. L. Gall, R. P. McAndrew, K. Deng, K. C. Holland, T. J. Donohue, D. R. Noguera, B. A. Simmons, K. L. Sale, et al., *J. Biol. Chem.* **2016**, *291*, 10228–38.

[72] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, *BMC Bioinformatics* **2009**, *10*, 1–9.

[73] P. K. Busk, L. Lange, *Appl. Environ. Microbiol.* **2013**, *79*, 3380–3391.

[74] Y. Huang, P. K. Busk, M. N. Grell, H. Zhao, L. Lange, *Enzyme Microb. Technol.* **2014**, *67*, 47–52.

[75] P. K. Busk, L. Lange, *BMC Genomics* **2015**, *16*, 1–13.

[76] P. K. Busk, B. Pilgaard, M. J. Lezyk, A. S. Meyer, L. Lange, *BMC Bioinformatics* **2017**, *18*, 1–9.

[77] K. R. Albe, M. H. Butler, B. E. Wright, *J. Theor. Biol.* **1990**, *143*, 163–195.

[78] P. K. Robinson, *Essays Biochem.* **2015**, *59*, 1–41.

[79] B. Sarrouh, T. M. Santos, A. Miyoshi, R. Dias, V. Azevedo, *J. Bioprocess. Biotech.* **2012**, *1*, 1–10.

[80] F. H. Arnold, A. A. Volkov, *Curr. Opin. Chem. Biol.* **1999**, *3*, 54–59.

[81] R. A. Sheldon, *Pure Appl. Chem.* **2000**, *72*, 1233–1246.

[82] K. Chen, F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 5618–22.

[83] D. S. Wilson, A. D. Keefe, in *Curr. Protoc. Mol. Biol.*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2001**.

[84] W. P. C. Stemmer, *Nature* **1994**, *370*, 389–391.

[85] J. Cohen, *Science* **2001**, *293*, 1517–1520.

[86] B. J. Jones, H. Y. Lim, J. Huang, R. J. Kazlauskas, *Biochemistry* **2017**, *56*, 6521–6532.

[87] Agilent Technologies, "QuikChange Site-Directed Mutagenesis Kit Instruction Manual," can be found under https://www.agilent.com/cs/library/usermanuals/public/200518.pdf, **2015**.

[88] NEB, "Q5 Site-Directed Mutagenesis Kit - New England Biolabs GmbH," can be found under https://www.neb-online.de/pcr-dna-amplifikation/q5-site-directed-mutagenesis-kit/, **2019**.

[89] M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, *Angew. Chemie Int. Ed.* **2005**, *44*, 4192–4196.

[90] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891–903.

[91] B. A. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., *Proteomics Protoc. Handbook, Humana Press* **2005**, 571–607.

[92] N. Kausar, P. P. Ghosh, G. Pal, A. R. Das, *RSC Adv.* **2015**, *5*, 60199–60207.

[93] M. Mirza-Aghayan, R. Boukherroub, M. Bolourtchian, M. Rahimifard, *J. Organomet. Chem.* **2007**, *692*, 5113–5116.

[94] P. Das, A. K. Mandal, U. Reddy G., M. Baidya, S. K. Ghosh, A. Das, *Org. Biomol. Chem.* **2013**, *11*, 6604–6614.

[95] A. Löytynoja, N. Goldman, *BMC Bioinformatics* **2010**, *11*, 579.

[96] J. Trifinopoulos, L.-T. Nguyen, A. von Haeseler, B. Q. Minh, *Nucleic Acids Res.* **2016**, *44*, 232–235.

[97] K. Katoh, J. Rozewicki, K. D. Yamada, *Brief. Bioinform.* **2017**, *20*, 1160–1166.

[98] G. Crooks, G. Hon, J. Chandonia, S. Brenner, *Genome Res* **2004**, *14*, 1188–1190.

[99] O. . O. A. J. Trott, *J.Comput.Chem.* **2010**, *31*, 455–461.

[100] X. Wang, K. Zheng, H. Zheng, H. Nie, Z. Yang, L. Tang, *J. Biotechnol.* **2014**, *192*, 102–107.

[101] K. Faber, W. Kroutil, "A Computer Program for the Determination of the Enantioselectivity (E-Value) in the Kinetic Resolution of Enantiomers," can be found under http://biocatalysis.uni-graz.at/enantio/DataFiles/Selectivity-Help.pdf, **2012**.

[102] C. S. Chen, Y. Fujimoto, G. Girdaukas, C. J. Sih, *J. Am. Chem. Soc.* **1982**, *104*, 7294–7299.

[103] I. Rahman, A. Kode, S. K. Biswas, *Nat. Protoc.* **2007**, *1*, 3159–3165.

[104] L. Yi, H. Li, L. Sun, L. Liu, C. Zhang, Z. Xi, *Angew. Chemie Int. Ed.* **2009**, *48*, 4034–4037.

[105] L. E. . Barata, L. S. Santos, P. H. Ferri, J. D. Phillipson, A. Paine, S. L. Croft, *Phytochemistry* **2000**, *55*, 589–595.

[106] J. Lu, Q. Song, Z. Ji, X. Liu, T. Wang, Q. Kang, *Electron. J. Biotechnol.* **2015**, *18*, 281–285.

[107] N. K. Tripathi, K. Sathyaseelan, A. M. Jana, P. V. L. Rao, *Def. Sci. J.* **2009**, *59*, 137–146.

[108] P. T. Anastas, J. M. Woodley, *Handb. Green Chem.* **2018**, *11*, 207–230.

[109] R. Wohlgemuth, *Chem. Biochem. Eng. Q.* **2017**, *31*, 131–138.

[110] W. a Greenberg, A. Varvak, S. R. Hanson, K. Wong, H. Huang, P. Chen, M. J. Burk, *Proc. Natl. Acad. Sci.* **2004**, *101*, 5788–5793.

[111] J. Husarcikova, A. Schallmey, *J. Biotechnol.* **2019**, *293*, 1–7.

[112] A. Basso, S. Cantone, C. Ebert, P. J. Halling, L. Gardossi, *Org. Synth. with Enzym. Non-Aqueous Media* **2008**,

279–301.

[113] R. H. Kimball, *J. Am. Chem. Soc.* **1936**, *58*, 1963–1970.

[114] N. Allocati, M. Masulli, M. Pietracupa, L. Federici, C. Di Ilio, *Biochem. J.* **2006**, *394*, 11–17.

[115] M. Widersten, R. H. Kolm, R. Bjornestedt, B. Mannervik, *Biochem. J.* **1992**, *285*, 377–381.

[116] M. Schallmey, J. Koopmeiners, E. Wells, R. Wardenga, A. Schallmey, *Appl. Environ. Microbiol.* **2014**, *80*, 7303–7315.

[117] S. Palamuru, N. Dellas, S. L. Pearce, A. C. Warden, J. G. Oakeshott, G. Pandey, *Appl. Environ. Microbiol.* **2015**, *81*, 8164–8176.

[118] X.-Y. Yu, X.-D. Yu, G.-Y. Fu, Z. Zhao, X. Shen, C. Sun, M. Wu, *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 2061–2067.

[119] D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, *Genome Res.* **2015**, *25*, 1043–1055.

[120] H. Xiao, Z. Bao, H. Zhao, *Ind. Eng. Chem. Res.* **2015**, *54*, 4011–4020.

[121] M. T. Reetz, K. M. Kühling, S. Wilensek, H. Husmann, U. W. Häusig, M. Hermes, *Catal. Today* **2001**, *67*, 389–396.

[122] M. Dörr, M. P. C. Fibinger, D. Last, S. Schmidt, J. Santos-Aberturas, D. Böttcher, A. Hummel, C. Vickers, M. Voss, U. T. Bornscheuer, *Biotechnol. Bioeng.* **2016**, *113*, 1421–1432.

Abbreviation

2,6-MP-VG	β-(2,6-dimethoxyphenoxy)-α-veratrylglycerone	IR	infrared radiation
2-MTHF	2-methyl-tetrahydrofuran	ISM	iterative saturation mutagenesis
2-MTP-VG	2-(2,6-dimethoxyphenoxy)-1-(3,4-dimethoxyphenyl)-3-hydroxypropan-1-one	KDL reaction	kinase DpnI ligase reaction
3,5-MP-VG	β-(3,5-dimethoxyphenoxy)-α-veratrylglycerone	Kpi	potassium phosphate buffer
AA	amino acid	LB media	lysogeny broth
ACN	acetonitrile	LC-MS	liquid chromatography–mass spectrometry
APS	ammonium persulfate	LED	light-emitting diode
ATP	adenosine triphosphate	LiPs	lignin peroxidases
BLAST	Basic Local Alignment Search Tool	MeoH	methanol
BLASTP	Protein BLAST	MIBK	methyl isobutyl ketone
B-PER	bacterial protein extraction reagent	MilliQ	ultrapure water
CEL	cellulolytic enzyme lignin	MnPs	manganese peroxidases
CHES	*N*-cyclohexyl-2-aminoethanesulfonic acid	MST	micro scale electrophoresis
CV	column volumes	MTBE	methyl *tert*-butyl ether
CYP450	cytochrome P450	MU-VG	β-(methylumbelliferyl)-α-veratrylglycerone
D. squalens	*Dichomitus squalens*	*N. aromaticivorans*	*Novosphingobium aromaticivorans*
DCM	dichloromethane	NAD	nicotinamide adenine dinucleotide
DDQ	2,3-dichloro-5,6-dicyano-1,4-benzoquinone	NADPH	nicotinamide adenine dinucleotide phosphate
DFT calculations	density functional theory	NBS	*N*-bromsuccinimid
DIPEA	*N*,*N*-diisopropylethylamine	NMR	nuclear magnetic resonance
DMF	dimethylformamide	NTA	nitrilotriacetic acid
DMSO	dimethyl sulfoxide	PCR	polymerase chain reaction
DNA	deoxyribonucleic acid	PDB	Protein Data Bank
DTNB	5,5'-dithiobis-(2-nitrobenzoic acid)	PEG	polyethylene glycol
E. coli	*Escherichia coli*	pI	isoelectric point
EMAL	enzymatic mild acidolysis lignin	PMSF	phenylmethylsulfonyl fluoride
EPA	environmental protection agency	PPR	peptide pattern recognition
ESI-HRMS	electrospray ionization high resolution mass spectrometry	qPCR	real-time polymerase chain reaction
ESI-MS	electrospray ionization mass spectrometry	*rac*	racemic
et al.	et alii (and others)	reac.	reaction
EtOAc	ethyl acetate	Rel.	relative
EtOH	ethanol	RMSD	root-mean-square deviation

FDCA	2,5-furandicarboxylic acid	RV	reverse
VN-VG	β-vanillinyl-α-veratrylglycerone	S	sinapyl alcohol
FW	forward	*S. xenophagum*	*Sphingobium xenophagum*
G	coniferyl alcohol	SDS	sodium dodecyl sulfate
GAV	β-(guaiacyl)-α-acetovanillone	SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
GGE	β-(guaiacyl)-α-guaiacylglycerol	SEC	size-exclusion chromatography
GPC	gel permeation chromatography	SOC	super optimal broth media
GR	glutathione reductase	SSM	site directed saturation mutagenesis
GS-AV	β-glutathionyl-α-acetoveratrone	*T. denitrificans*	*Thiobacillus denitrificans*
GSH	glutathione	*T. versicolor*	*Tinea Versicolor*
GSSG	oxidized glutathione	TAE	Tris base, acetic acid, EDTA buffer
GST	glutathione S-transferase	TB media	terrific broth
GS-VG	β-glutathionyl-α-veratrylglycerone	TEMED	tetramethylethylendiamine
GVE	β-(guaiacyl)-α-veratrylethanone	Temp.	temperature
GVG	β-(guaiacyl)-α-veratrylglycerone	TEMPO	2,2,6,6-tetramethyl-1-piperidinyloxy
GVL	β-guaiacyl-α-veratrylglycerol	THF	tetrahydrofuran
H	*p*-coumaryl alcohol	T_m	melting temperature
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	TOF	time of flight
HPLC	high performance liquid chromatography	Tris	tris(hydroxymethyl)aminomethane
HSP	high-scoring segment pair	UV-Vis	Ultraviolet–visible spectroscopy
HTS	high-throughput screening	VA	veratryl alcohol
IMAC	immobilized ion affinity chromatography	VG	β-deoxy-α-veratrylglycerone
IPTG	isopropyl-β-D-thiogalactopyranoside	wt%	weight %

List of figures

List of tables

Acknowledgements

My special thanks go to Prof. Dr. Anett Schallmey for enabling my PhD-thesis in her work group, as well as for accompany this process throughout the whole time. I really enjoyed the time and was always very satisfied with the support. I am also very grateful for the freedom of creativity, which offered to me during the individual projects. I would also like to thank for participating of in total five international conferences as well as two scientific consortia (Carbazymes, PROCOMPAS), something Anett has actively supported.

Prof. Dr. Dieter Jahn and Prof. Dr. Ralf Mendel I would like to thank for being part in my examination board, Dieter Jahn as second reviewer and Ralf Mendel as chairman. I know them both for several years, especially through the graduate college PROCOMPAS, where I was also allowed to participate. I thank Dieter Jahn and Ralf Mendel also for the organization of this graduate college.

The project of my doctoral thesis was financed within the MWK project by the state of Lower Saxony, also I would like to thank for this.

This PhD-thesis would not have been possible without various cooperation partners. At the start, but also during the whole time, the Inorganic Chemistry of Prof. Dr. M. Bröring and the microbiology of Prof. Dr. Dieter Jahn supported my work with laboratory space, laboratory equipment, analytics and chemicals. I would especially like to thank Simone, Jörn, Stefanie and Julian from these working groups for their help.
I would like to thank the protein crystallography work group of Prof. Dr. Wulf Blankenfeld for various crystallization attempts including the associated purification as well as for the successful solution of some crystal structures. Especially the employees Christina and Thomans have to be mentioned here.
Theoretical Chemistry by Prof. Dr. Christoph Jacob and his coworker Julian did DFT calculations for me. I would also like to thank the Chemical Biology of Prof. Dr. Mark Brönstrup and Ulrike for LC-MS analyzes.

For the long and successful collaboration, I thank all employees of the Biochemistry Working Group. I would like to thank Marcus for the working together on the "Etherases Publication". Melanie, is the good soul of the working group, has always an open ear and is trying to help with all organizational matters. From Patrick I learned a lot about docking. With Kathi I always liked to discuss about chemical syntheses, and I would also like to thank Kathi for corrections of my thesis. Hazel showed me the primer design for Q5-mutagenesis, which greatly supported my protein engineering experiments. Willem was a great help in many situations, due to his vast biochemical expertise. At the beginning Julia introduced me to the standard working methods of the working group and also corrected parts of my thesis. Together with Jhon Alex I was part of PROCOMPAS, also through him this was a very good experience. Jennifer has always been a very helpful colleague and helped me by GC measurements. Rita I

will not forget cause of her honest sense of humor and the shared time together in the Carbazymes project, where we fought our way. Jana worked with me on the lignin theme. Therefore, I would like specially to thank Jana for the good cooperation. For the fact that we always helped each other and always found a common solution.

I would also like to thank my students, whom I have supervised during bachelor theses and research internships. Carina, Patrick, Rita, Lars, and Wiebke have done a great job and played a big part in this work.

My parents I thank for the great support. They have enabled and funded my studies and through their education, made me what I am today.

Sandra stays with me for 8 years now, most of my studies, the whole PhD and hopefully far beyond. Without Sandra I probably would not have started my PhD in Braunschweig neither managed it. In addition to the support for the work, I am also very grateful for the whole time together, like climbing or geocaching, and looking forward to the future.

Mein besonderer Dank gilt Prof. Dr. Anett Schallmey für das ermöglichen meiner Promotion in Ihrer Arbeitsgruppe, als auch für die Begleitung dieses Prozesses während der ganzen Zeit. Ich habe die gesamte Promotion sehr genossen und auch mit der Betreuung war ich immer sehr zufrieden. Sehr dankbar bin ich auch für die Gestaltungsmöglichkeiten, die mir während der einzelnen Projekte ermöglicht wurden. Zudem möchte ich mich für die Möglichkeit der Teilnahme an insgesamt fünf internationalen Konferenzen sowie an zwei wissenschaftlichen Konsortien (Carbazymes, PROCOMPAS) bedanken, etwas was Anett auch aktiv gefördert hat.

Bei Herrn Prof. Dr. Dieter Jahn und Herrn Prof. Dr. Ralf Mendel möchte ich mich herzlich bedanken für die Beteiligung an meiner Prüfungskommission, Herr Jahn als Zweitgutachter und Herr Mendel als Vorsitzender. Beide kenn ich schon seit einigen Jahren, besonders durch das Graduiertenkollegs PROCOMPAS, an dem ich auch teilnehmen durfte. Auch für die Organisation des Kollegs möchte ich mich bei Herrn Jahn und Herrn Mendel bedanken.

Das Projekt meiner Doktorarbeit wurde im Zuge des MWK-Projektes vom Land Niedersachen finanziert, auch hierfür möchte ich mich bedanken.

Diese Doktorarbeit wäre auch ohne diverse Kooperationspartner nicht möglich gewesen. Am Start, aber auch während der gesamten Zeit, haben die Anorganische Chemie von Prof. Dr. M. Bröring und die Mikrobiologie von Prof. Dr. Dieter Jahn mit Laborplatz, Laborgeräten, Analytik und Chemikalien meine Arbeit unterstützt. Besonders danken möchte ich Simone, Jörn, Stefanie und Julian aus diesen Arbeitsgruppen für Ihre Hilfe.
Der Proteinkristallographie von Prof. Dr. Wulf Blankenfeld möchte ich für diverse Kristallisationsversuche inklusive der dazugehörigen Aufreinigung sowie das erfolgreiche lösen einiger Kristall-strukturen danken. Hier sind besonders die Mitarbeitern Christina und Thomans zu erwähnen.
Die Theoretische Chemie von Prof Dr. Christoph Jacob mit dem Mitarbeiter Julian hat für mich DFT-Kalkulationen durchgeführt. Auch der Chemischen Biologie von Prof. Dr. Mark Brönstrup und Ulrike möchte ich für LC-MS Analysen danken.

Für die lange und erfolgreiche Zusammenarbeit möchte ich auch allen Mitarbeitern der Biochemie Arbeitsgruppe danken. Marcus möchte ich für die gemeinsame Arbeit an der „Etherasen Publikation“ danken. Melanie ist die gute Seele der Arbeitsgruppe, immer mit einem offenen Ohr und versucht bei allen organisatorischen Dingen zu helfen. Von Patrick Schrepfer habe ich sehr viel über Dockings gelernt. Mit Kathi habe ich immer gerne über chemische Synthesen diskutiert und ich möchte Kathi auch für Korrekturen meiner Dissertation danken. Hazel hat mir das Primerdesign für die Q5-Mutagenese gezeigt, was meine Protein engineering-Experimente sehr vereinfacht hat. Willem war durch sein gewaltiges biochemisches Fachwissen in vielen Situationen eine sehr große Hilfe. Julia hat mich zu Beginn in die Standardarbeitsweisen der Arbeitsgruppe eingewiesen und auch Teile der Dissertation korrigiert. Zusammen mit Jhon Alex habe ich an PROCOMPAS teilgenommen, auch durch ihn war dies eine sehr gute Erfahrung. Jennifer war immer eine sehr hilfsbereite

Kollegin und hat mir auch bei GC-Messungen geholfen. Rita werde ich vor allem durch ihren ehrlichen Humor als auch durch die gemeinsame Zeit beim Carbazymes-Projekt, wo wir uns gemeinsam durchgekämpft haben, in Erinnerung behalten. Jana hat mit mir gemeinsam das Ligninthema bearbeitet. Daher möchte ich Jana auch besonders für die gute Zusammenarbeit danken. Dafür das wir einander immer geholfen haben und immer eine gemeinsame Lösung gefunden haben.

Ich möchte mich auch bei meinen Studenten, die ich im Zuge von Bachelorarbeiten und Forschungspraktika betreut haben, bedanken. Carina, Patrick, Rita, Lars, Wiebke haben großartige Arbeit geleistet und einen großen Anteil an dieser Arbeit.

Meinen Eltern danke ich für die großartige Unterstützung, sie haben mir das Studium ermöglicht und finanziert und mich durch ihre Erziehung zu dem gemacht was ich heute bin.

Sandra begleitet mich inzwischen seit 8 Jahren, ein Großteil meines Studiums, die ganze Promotion und hoffentlich noch weit darüber hinaus. Ohne Sandra hätte ich wohl weder meine Promotion in Braunschweig angefangen noch diese geschafft. Neben der Unterstützung für die Arbeit, bin ich auch für die gesamte gemeinsame Zeit, wie beim gemeinsamen klettern oder geocachen, sehr dankbar und freue mich auf die Zukunft.

Student works and internships

Carina, Bachelor thesis “Klonierung, Reinigung und Charakterisierung neuer β-Etherasen“ in **2017**, cloning of LigE889, LigF935, LigG817, and GST3 as well as characterization of LigE283, LigE889, LigF008 and LigF935. → **section 3.3**

Lars, research internship “β-etherases and glutathione-lyases for selective lignin depolymerization” in **2017**, characterization of LigG817 and GST3 → **section 3.5.1**

Rita, CB13 Biokatalyse internship in **2017**, generation and characterization of LigG-TD mutants Y112A and Y112W → **section 3.4.4**

Wiebke, Research Internship “Investigation of an assay for the prescreening of gluthathione dependant lyases for lignin depolymerization and application on a SSM library of LigG2 mutants” in **2018**, development of glutathione reductase based HTS assay as well as generation and screening of LigG-TD A55 SSM library. → **section 3.4.3 and 3.4.5**

Patrick, Bachelor thesis “Protein engineering of a glutathione lyase and establishment of an assay for determination of relative activity” in **2018**, optimization of the glutathione reductase based HTS assay as well as generation of the LigG-TD SSM libraries V11, D104, R105, V108, M116, F165, and N223 and screening of LigG-TD SSM libraries A55, V11, D104, and F165. → **section 3.4.3 and 3.4.5**

Veröffentlichungen

Publikationen

M. Frese, P. H. Guzowska, H. Voß, N. Sewald, "*Regioselective Enzymatic Halogenation of Substituted Tryptophan Derivatives using the FAD-Dependent Halogenase RebH*", *CHEMCATCHEM*, Vol. 6, No. 5, 1270-1276, (**2014)**.

M. Frese, C. Schnepel, H. Minges, H. Voß, R. Feiner, N. Sewald, "*Modular Combination of Enzymatic Halogenation of Tryptophan with Suzuki–Miyaura Cross-Coupling Reactions*", *CHEMCATCHEM*, Vol. 8, No. 10, 1799-1803, (**2016)**.

J. Koopmeiners, C. Diederich, J. Solarczek, H. Voß, W. Blankenfeld and A. Schallmey, "HheG, a Halohydrin Dehalogenase with Activity on Cyclic Epoxides", ACS Catal., Vol. 7, No. 10, 6877-6886, (**2017**).

J. Husarcíková, H. Voß, P. Domínguez de María, A. Schallmey, "*Microbial* β-*etherases and glutathione lyases for lignin valorisationin biorefineries: current state and future perspectives*", *Applied Microbiology and Biotechnology*, Vol 102, 5391–5401, (**2018**)**.**

H. Voß, C. A. Heck, M. Schallmey, A. Schallmey, "*Database mining for novel bacterial* β-*etherases, glutathione-dependent lignin-degrading enzymes*", Applied and Environmental Microbiology, Vol 86, No 2, (**2020**).

Posterbeiträge

H. Voß, P. Picart, C. Diederich, A. Schallmey, "*Application of* β-*etherases and glutathione lyases in lignin valorization*", **Biocat 2016**, 8th International Congress of Biocatalysis, Hamburg, Germany, (**2016**).

H. Voß, C.Diederich, A. Schallmey, "*PROTEIN ENGINEERING OF LIG-TD, A GLUTATHTION LYASE***", BioTrans 2017**, 13th International Symposium on Biocatalysis and Biotransformation, Budapest, Hungary, (**2017**).

H. Voß, C. A. Heck, A. Schallmey, "*IDENTIFICATION OF NEW* β-*ETHERASES BY PEPTIDE PATTERN RECOGNITION*", **ICRC 2018**, International CeBiTec Research Conference 2018 Reaction Concepts for Industrial Biocatalysis, Bielefeld, Germany, (**2018**).

H. Voß, C. A. Heck, A. Schallmey, "*IDENTIFICATION OF NEW* β-*ETHERASES BY PEPTIDE PATTERN RECOGNITION*", **Biocat 2018**, 9th International Congress of Biocatalysis, Hamburg, Germany, (**2018**).

H. Voß, W. P. Dijkman, J. Wessel, P. Schrepfer, S. Charnock, A. Schallmey, "*New ketoreductases for the selective oxidation of a statin side chain precursor*", **Novel Enzymes 2018**, 6th Internation Conference on Novel Enzymes, Darmstadt, Germany, (**2018**).

H. Voß, C. A. Heck, A. Schallmey, "*IDENTIFICATION OF NEW* β-*ETHERASES BY PEPTIDE PATTERN RECOGNITION*", **Procompas PhD Symposium 2018**, Braunschweig, Germany, (**2018**).